V 465

13 DEC 71

D-5

11 2123

NAVAL RESEARCH LOGISTICS QUARTERLY

JUNE 1971
VOL. 18, NO. 2

OFFICE OF NAVAL RESEARCH

NAVSO P-1278

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

H. E. Eccles
Rear Admiral, USN (Retired)

O. Morgenstern
New York University

F. D. Rigby
Texas Technological College

D. M. Gilford
U.S. Office of Education

S. M. Selig
Managing Editor
Office of Naval Research
Arlington, Va. 22217

ASSOCIATE EDITORS

R. Bellman, RAND Corporation
J. C. Busby, Jr., Captain, SC, USN (Retired)
W. W. Cooper, Carnegie Mellon University
J. G. Dean, Captain, SC, USN
G. Dyer, Vice Admiral, USN (Retired)
P. L. Folsom, Captain, USN (Retired)
M. A. Geisler, RAND Corporation
A. J. Hoffman, International Business Machines Corporation
H. P. Jones, Commander, SC, USN (Retired)
S. Karlin, Stanford University
H. W. Kuhn, Princeton University
J. Laderman, Office of Naval Research
R. J. Lundegard, Office of Naval Research
W. H. Marlow, The George Washington University
B. J. McDonald, Office of Naval Research
R. E. McShane, Vice Admiral, USN (Retired)
W. F. Millson, Captain, SC, USN
H. D. Moore, Captain, SC, USN (Retired)
M. I. Rosenberg, Captain, USN (Retired)
D. Rosenblatt, National Bureau of Standards
J. V. Rosapepe, Commander, SC, USN (Retired)
T. L. Saaty, University of Pennsylvania
E. K. Scofield, Captain, SC, USN (Retired)
M. W. Shelly, University of Kansas
J. R. Simpson, Office of Naval Research
J. S. Skoczylas, Colonel, USMC
S. R. Smith, Naval Research Laboratory
H. Solomon, The George Washington University
I. Stakgold, Northwestern University
E. D. Stanley, Jr., Rear Admiral, USN (Retired)
C. Stein, Jr., Captain, SC, USN (Retired)
R. M. Thrall, Rice University
T. C. Varley, Office of Naval Research
C. B. Tompkins, University of California
J. F. Tynan, Commander, SC, USN (Retired)
J. D. Wilkes, Department of Defense OASD (ISA)

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Information for Contributors is indicated on inside back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Subscription Price: $5.50 a year in the U.S. and Canada, $7.00 elsewhere. Cost of individual issues may be obtained from the Superintendent of Documents.

The views and opinions expressed in this quarterly are those of the authors and not necessarily those of the Office of Naval Research.

Issuance of this periodical approved in accordance with Department of the Navy Publications and Printing Regulations, NAVEXOS P-35

Permission has been granted to use the copyrighted material appearing in this publication.

OPERATIONS ANALYSIS DURING THE UNDERWATER SEARCH FOR SCORPION*

Henry R. Richardson
and
Lawrence D. Stone

Daniel H. Wagner, Associates
Paoli, Pennsylvania

ABSTRACT

This paper discusses the operations analysis in the underwater search for the remains of the submarine *Scorpion*.

The a priori target location probability distribution for the search was obtained by monte-carlo procedures based upon nine different scenarios concerning the *Scorpion* loss and associated credibility weights. These scenarios and weights were postulated by others. *Scorpion* was found within 260 yards of the search grid cell having the largest a priori probability.

Frequent computations of local effectiveness probabilities (*LEP*s) were carried out on scene during the search and were used to determine an updated (a posteriori) target location distribution. This distribution formed the basis for recommendation of the current high probability areas for search.

The sum of *LEP*s weighted by the a priori target location probabilities is called search effectiveness probability (*SEP*) and was used as the overall measure of effectiveness for the operation. *SEP* and *LEP*s were used previously in the Mediterranean H-bomb search.

On-scene and stateside operations analysis are discussed and the progress of the search is indicated by values of *SEP* for various periods during the operation.

This paper discusses the operations analysis effort made by the authors' firm, on scene and stateside, during the search for the lost submarine *Scorpion*, June–October 1968. We present an outline of the methods employed and advice given during the search, as well as the impact of the operations analysis on the conduct of the search. An appraisal of the analysis effort is also presented. Because of estimates made of sensor capabilities after *Scorpion* was found, we are able to present a review of the effectiveness of the search effort as it developed in time using the after-search estimates of sensor capabilities and compare these estimates of effectiveness to those made during the search.

In general, the operations analysis provided a quantitative framework for the planning and documentation of the search effort; of itself this substantially enhanced the continuity of the search in the face of considerable turnover of leadership and other personnel. The computation of the a priori distribution of target location proved to be highly successful both as a predictor of location and as a basis for estimation of total required search effort to be expected. The objective of providing good guidance for allocation of search effort was reasonably achieved, although various factors, principally overestimation of sensor capabilities, detracted from the quality of this guidance. The single substantial

*The analysis reported here was largely performed by Daniel H. Wagner, Associates under Ocean Systems, Incorporated Purchase Order No. 863-3539 as a subcontract to Contract No. N00024-69-C-0201 with the Supervisor of Salvage, Naval Ship Systems Command.

contact investigation effort was guided by operations analysis, which also provided a basis for concluding this effort.

The first section provides general background on the operation. Sensors, navigation, the search grid, and the a priori distribution are discussed. The primary measure of effectiveness used, search effectiveness probability (*SEP*), is mentioned with a more detailed explanation included in appendix A. The next two sections discuss separately the operations analysis as it actually took place on scene and stateside. The on-scene analysis of the investigation of contact M8/3 is given in appendix B. After-search estimates of search effectiveness are presented in the fourth section. These values are based upon estimates of sensor sweep width obtained using data collected on scene after *Scorpion* was found. An appraisal of the operations analysis effort is given in the final section.

General Background

After the emergency search phase which stretched across the Atlantic Ocean, the search effort was confined to an area located approximately 400 miles southwest of the Azores.

The search operation was under the tactical command of a succession of six commodores, each an Atlantic Fleet submarine squadron commander. Most of the search in this area was conducted by the *USNS Mizar* with which the embarked ocean engineering team from NRL has had a continuing working relationship; the instruments on board *Mizar* were manned by NRL and Navy Oceanographic Office personnel with help from industry personnel who were on board until August 14. The scientific personnel were divided into two shifts each of which worked 12 hours a day allowing the search to continue around the clock 7 days a week.

The search phases can be conveniently described in terms of *Mizar's* five cruises: (1) June 10–28, (2) July 10–August 6, (3) August 14–September 8, (4) September 20–October 7, and (5) October 16–November 2. These are dates out of the Azores which were 1.5 days sailing from the search scene.

On-scene operations analysis was supplied by the authors (Richardson on cruise (1) and Stone on the second half of cruise (2) and the first half of cruise (3)) and by student colleagues of theirs (S. G. Simpson, a doctoral candidate in mathematics at MIT, on the latter half of cruise (3) and J. A. Rosenberg, a Drexel University cooperative student, cruises (4) and (5)). During cruises (4) and (5), primary responsibility for operations analysis was assigned to Lt David Brummersted, USN and Lcdr James Finlen, USN of Submarine Development Group One (which has a deep submergence mission) who had been briefed on the main techniques presented herein—they were assisted by Rosenberg.

Great difficulties were involved in this search because of the depth of the ocean, the remoteness of this region, and, most of all, the lack of direct information as to the location of *Scorpion*. Despite these problems, *Scorpion* was located on October 28.

Search sensors. The search equipment consisted of a towed platform (called the sled) upon which were mounted cameras, magnetometers, and sonars. This platform operated as much as 2 miles below the surface of the ocean, creating difficult maneuverability problems.

The visual sensors consisted of wide angle lens cameras some of which could be loaded with enough film to take pictures every 30 seconds for as long as 30 hours.

The magnetometer which was employed throughout the search was a proton precession type which measured the strength, but not the direction of the magnetic field. It, too, was activated by a signal from *Mizar*. Despite the uneven magnetic background in the search area, the magnetometer was the first sensor to detect *Scorpion*.

Two different side-looking sonars were used during the search. One of these could be used in either short range or long range mode; the short range mode providing the higher resolution. This was the sonar which was used at the beginning of the search and was replaced by the second sonar after August 4. The latter sonar operated in a single mode.

Navigation. Because the search area was 400 miles away from the nearest land, there was no useful land-based navigation system available. *Loran* and *Omega* although available were not precise enough for conducting this type of search. Thus, navigation throughout the search was done with the aide of satellite fixes available about 18 times a day at irregular intervals and with respect to transponders which were anchored to the ocean bottom. The Navy Oceanographic Office provided the equipment and personnel to obtain the satellite fixes. Inertial navigation was *not* available; this degraded the accuracy usually ascribed to the satellite navigation system.

During most of the search, navigation was performed with uncoded transponders. This meant that one had to guess which transponder or transponders were responding at a given time. This caused considerable confusion and introduced large navigational errors at times.

When coded transponders were available, a computer-transponder system could be used to calculate the position of the ship and sled. This system consisted of coded transponders anchored to the bottom, three transceivers mounted in the hull of *Mizar*, and a responder located on the sled which carried the search sensors. The coding of the transponders allowed *Mizar* to interrogate a particular transponder without interrogating the entire field of them.

Estimates made on scene indicated that the error in this tracking system had a standard deviation of approximately 300 feet in any direction. This was a considerable improvement over the other navigational system using satellites and uncoded transponders.

Search grid. In order to allocate search effort effectively, the search grid shown in Figure 1 was established. The cells are 1 mile in the north-south direction and 0.84 miles in the east-west direction. This figure indicates several points of interest to the analysis. These are:

(i) Location of the piece of metal which was found early in the search and which was subsequently determined to be part of the *Scorpion* debris.

(ii) Sonar contact—This was a contact found in June; however, the area in which this contact is located is quite mountainous.

(iii) M8/3—The location of a large magnetic anomaly found on August 3. This anomaly was investigated quite thoroughly and determined to be a large magnetic rock.

(iv) The location of the largest piece of the *Scorpion*, which was photographed on October 28. For convenience, we index these rectangles either by the integers $j = 1, \ldots, N$, or by a letter-numeral combination (e.g., C8).

A priori distribution. Preparation of the a priori distribution (see Figure 2) superseded earlier attempts to define the search area.

As soon as certain preliminary inquiries and investigations into the possible locations of *Scorpion* were completed (by others), a conference was held July 18 where Dr. J. P. Craven and Dr. F. A. Andrews postulated nine scenarios concerning the events attending the disaster and assigned credences to them. These were then converted by one of the authors to individual a priori probability distributions of the location of *Scorpion* corresponding to the respective scenarios, and these were combined as an average weighted by the assigned credences to obtain the overall distribution (see Figure 2). Craven had earlier

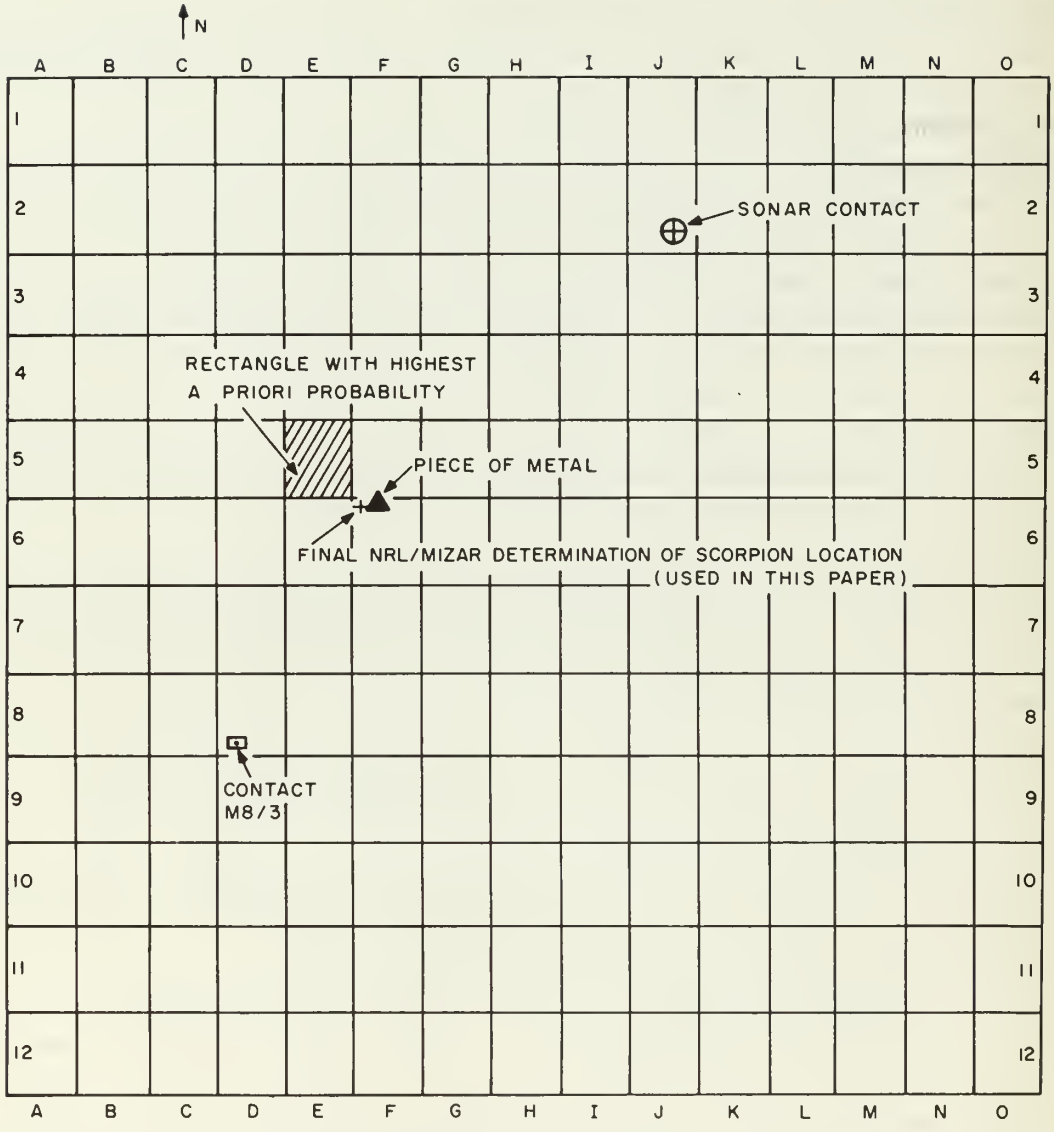

N
A B C D E F G H I J K L M N O
1 2 3 4 5 6 7 8 9 10 11 12
SONAR CONTACT
RECTANGLE WITH HIGHEST
A PRIORI PROBABILITY
PIECE OF METAL
FINAL NRL/MIZAR DETERMINATION OF SCORPION LOCATION
(USED IN THIS PAPER)
CONTACT
M8/3

FIGURE 1. *Scorpion* search grid

originated this approach in the H-bomb search. The calculations were performed by monte-carlo procedures.

The credence (a number between zero and one) associated with each scenario reflected the scenario's plausibility relative to the others and incorporated (at least indirectly as interpreted by Craven and Andrews) the views and opinions of Navy operating personnel as well as the analysis of specialists in many scientific areas.

The monte-carlo computer program for generating the probability map begins a replication by causing a random number to be drawn in order to determine the choice of scenario. The scenarios are selected with frequency specified by the assigned credences.

The movement of the submarine as specified by the selected scenario is then simulated with random numbers drawn as required to represent the uncertainties in course, speed, and position,

at the time the emergency occurred, as well as other variables. The final position of *Scorpion* is determined and the output of the replication is a "one" added to the appropriate search cell. The result of 10,000 replications is shown in Figure 2.

It is interesting to note that two search cells, E5 and B7, have 23 percent probability of enclosing *Scorpion*. It has now been determined that *Scorpion's* location is in rectangle F6 within 260 yards of the edge of rectangle E5, the rectangle having distinctly the highest a priori probability. Considering the size of the area covered by the a priori distribution, this is indeed a near miss.

Because of the efficiency associated with long sweeps, the piece of metal found early in the search, and other reasons, search effort was never assigned *solely* to cell E5, even when it was the current (a posteriori) high probability cell. Rather, search effort was assigned to the general vicinity of E5 with large amounts placed in neighboring cells, including F6.

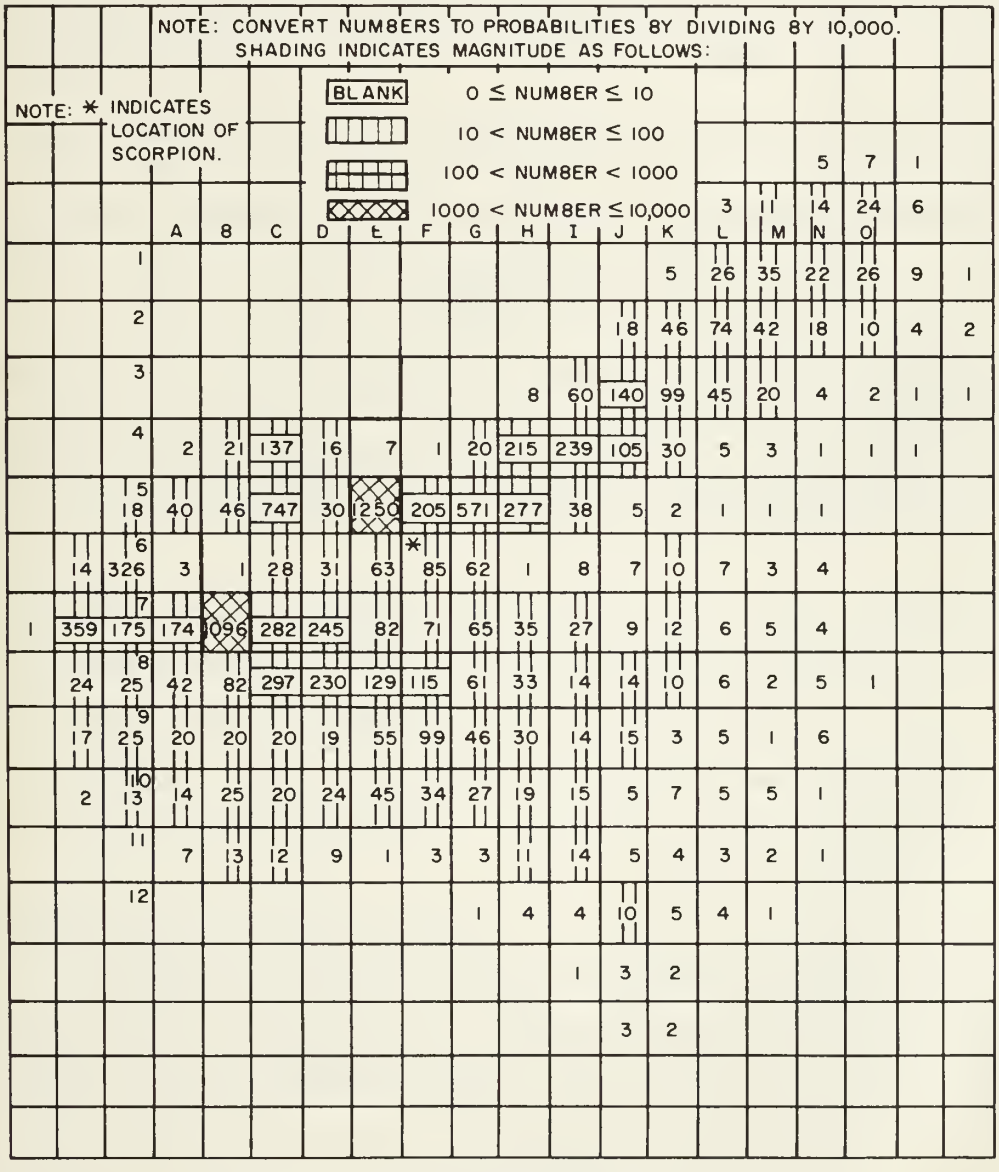


FIGURE 2. Overall A Priori distribution for *Scorpion* search

Calculations show, however, that even if effort had been confined to the search cells having highest a posteriori probability, then assuming a sweep width of 800 feet, the cells to be searched would have included F6 except for the first 153 miles of track length.

In the actual search, 1,026 miles of track were used for broad-area search (not including contact investigation); thus by comparison, 153 miles (6.4 days at 24 miles per day) does not seem too costly a delay in reaching the target's cell. Note that the sweep width of 800 feet assumed in the above calculation is actually smaller than the combined-sensor sweep width used in most of the search (see the second section to follow).

Measure of effectiveness. The measure of search effectiveness, *SEP*, used in this search is basically the same one as developed for the Mediterranean H-bomb search (see appendix A). Denote by L_j the a priori probability that *Scorpion* was in rectangle j and by E_j the probability that *Scorpion* would have been found and identified by the search effort *if Scorpion* were in rectangle j. The probability E_j is called local effectiveness probability (*LEP*), and search effectiveness probability (*SEP*) is defined as

$$\text{SEP} = \sum_{j=1}^{N} L_j E_j.$$

From the *LEP*'s and the a priori probabilities, the a posteriori probabilities L_j' are computed by

$$L_j' = \frac{L_j(1-E_j)}{1-SEP};$$

this is the probability that *Scorpion* is in rectangle j considering the (unsuccessful) search from which the E_j's are computed.

Incremental maximization of *SEP* was the criterion used to recommend search procedures. In practice, this criterion usually resulted in the recommendation that broad area search be conducted in the search cells having the largest values of L_j'.

Few contacts were logged during the search and only two were considered important. One of these, contact *M*8/3, received an extensive investigation (August 14 to August 26) which is discussed in appendix B. Incremental maximization of *SEP* was not being carried out during this investigation.

The computation of *LEP* and *SEP* requires a knowledge of the sensor capabilities, the navigational uncertainties of the search system, and the amount of track length contributed by the sensor groups (i.e., combinations of sensors used simultaneously from the same towed platform). As in the Palomares H-bomb search and many other searches, knowledge of the capabilities of the sensors against the search target was uncertain.

Throughout the analysis we characterize a sensor by two numbers M and β; M is twice the *maximum* detection range of the sensor, and β is an average detection probability of the sensor given that the sensor passes the target at lateral range r, averaged with respect to r over $[0, M]$.

More precisely, let $f(x)$ denote the probability of detecting the target, given that the distance of closest point of approach to the sensor moving on a straight track (the "lateral range") is x. The "sweep width" W (see Reference [1]) of the sensor is defined by

$$W = \int_{-\infty}^{\infty} f(x)dx,$$

and β is given by the expression

$$W = \beta M,$$

where M is defined above.

Often several sensors are towed on the same platform. The sweep Ω of the "sensor group" is computed by the formula

$$\Omega = W_n + \sum_{j=1}^{n-1} W_j \prod_{k=j+1}^{n} (1-\beta_k), \tag{1}$$

where the n individual sensors in the group are characterized by the parameters (M_j, β_j) for $1 \leq j \leq n$; the sensors are ordered so that $M_{j+1} \geq M_j$. Formula (1) is based on the assumption that each sensor on the platform operates independently. "Correlation" is introduced by the fact that on a given sweep, the target's lateral range is the same for all sensors in the group.

Operations Analysis On Scene

On-scene operations analysis was used to document the search effort and give day-to-day advice on the conduct of the search. To this end records of the track length in each search cell were kept for each sensor group in Table A–1 and *SEP*'s and posterior target location probabilities were computed regularly.

The operations analysis during the first cruise established a search grid and evaluated the effectiveness of the search effort up to that time. At this time no formal a priori distribution was available.

From June 27 to July 23 there was a hiatus in the on-scene analysis effort aboard *Mizar*. During this time, two contacts were found. A shiny, bent piece of metal was photographed, near the border of *F*5 and *F*6, and a large sonar contact, approximately three hundred feet long was found in *J*2. Both of these contacts received cursory investigations which were not successful in regaining or rephotographing the contacts. We now know that the piece of metal photographed at this time was part of the *Scorpion* debris.

When the on-scene analysis resumed on July 23, during the latter half of the second cruise, the computation of effectiveness was resumed. During run 28 on August 3, a large magnetic anomaly was found in rectangle *D*8. This anomaly was designated *M*8/3. A brief and apparently successful attempt was made to regain this contact on the magnetometer and then *Mizar* returned to port.

In the third cruise, it was decided to investigate and identify the contact *M*8/3. The operations analysis provided assistance in devising search plans for investigating this contact, in evaluating the effectiveness of the contact investigation, and in recommending the allocation of investigation effort.

The investigation of *M*8/3 was discontinued on August 26 after run 52. At this time, it was calculated that the probability that if *M*8/3 were *Scorpion*, then it would have been identified, was 0.86. The details of this analysis are given in appendix B.

During the investigation of *M*8/3, estimates were made of the camera sweep width and of the uncertainty in the computer-transponder navigation system. From records kept during the investigation of the sonar contact in *J*2, is was determined that the probability that if the sonar contact were *Scorpion*, then it would have been identified, was 0.06.

Some of the previous runs had produced photographs of small apparently man-made objects (called artifacts) which might have come from *Scorpion*. These and subsequent artifacts were carefully logged. It was decided to try to determine the size and density of this field of artifacts. Thus although

the a priori distribution had arrived on scene by this time and *LEP*'s and *SEP*'s were being computed by the on-scene analyst, these probabilistic considerations did not play a substantial part in determining the placement of search effort at this time.

In the fourth cruise, some effort was made to extend and further define the artifact field, however, this effort met with little success and was soon abandoned. During the remaining part of this phase, which lasted until October 9, the *LEP*'s played an important role in determining the allocation of search effort. This caused a significant amount of effort to be placed in the areas of high probability, notably in *E*5. One of the principal objectives of the search during this cruise was to incrementally maximize *SEP*; this usually resulted in recommendation to search in the cells of highest a posteriori probability. This procedure approximates the Koopman optimal allocation of search effort (Reference [1]) in the case where there are no false contacts. Recent research by the authors' firm has resulted in better search criteria when false contacts are present (see References [3] and [4]).

The fifth cruise was originally intended to be a research and development/search cruise to test sensor modifications, to refine the underwater tracking system, and to study bottom contours and attempt regular search patterns. Search effort was eventually applied to a region containing *F*6 (where *Scorpion* was found) because of proximity to the piece of metal and the fact that this was a region with high a priori probability. On run 74 the magnetometer registered several sharp anomalies. It was possible to return to the same region and on the next run to regain the original contacts and to photograph the *Scorpion* on October 28. On November 1 the on-scene analysis terminated with Rosenberg returning stateside as a courier of photographs of *Scorpion*.

Table 1 gives *SEP* as well as *LEP* in *F*6 as they were computed on scene for search terminating September 10, October 9, and October 28.

TABLE 1. *On-Scene SEP and LEP in F6*

	Runs			
	1–38[a]	1–63	1–71	1–74
	Termination Date			
	Aug. 6	Sept. 10	Oct. 9	Oct. 28
On-scene *SEP*	.....	0.40	0.73	0.75
On-scene *LEP* in *F*6	.....	0.65	0.94	0.99

[a] Individual sensor detection probabilities, rather than *LEP* and *SEP*, were computed on scene through run 38.

The *LEP*'s computed on scene for *F*6 are, to a surprising degree, higher than the after-search estimates (see Table 3 in the penultimate section), largely because of the optimistic sensor (particularly sonar) sweep width estimates then available to the analysts. Reference [2] and section 5 of [4] are addressed to optimal search with uncertain sweep width data and were motivated by problems arising in this search.

As noted *Scorpion* location is less than 260 yards from the edge of *E*5, which, despite the optimistic values of sweep width used, had the highest a posteriori probability, computed on scene, of containing *Scorpion* during most of the search. At the end of run 71, the last time on-scene a posteriori prob-

abilities were computed, only 2 percent of the rectangles had higher a posteriori probabilities than *E*5. Since parts of *Scorpion* were found close to *E*5, the a posteriori probabilities provided good advice as to where to search, despite again the adverse influence of unrealistic sonar sweep width estimates then available.

Operations Analysis Stateside During the Search

Although the long lines of communication impeded coordination between on-scene and stateside analysis, during the search useful supporting operations analyses were being done stateside. The most important of these was the calculation of the a priori distribution (see Figure 2) of the location of *Scorpion* from scenarios and scenario credences supplied by Craven and Andrews.

The a priori distribution along with *LEP*'s provided a basis for an orderly allocation of search effort by the method described in the first section. The a priori distribution was completed and the estimates of required search time based upon it were transmitted to the Technical Analysis Group on August 1. The distribution was then hand-carried at the earliest date to the scene of operations arriving August 12.

Thus, it was only during the at-sea periods beginning with the third cruise that the a priori distribution had an opportunity to play a role in the conduct of the search.

Even though the a priori distribution was not available on-scene for the first two months of the search effort, it should be noted that the presence of an a priori distribution during the latter part of a search coupled with estimates of effectiveness during the initial part of the search can be used to bring the probability of detection eventually to the level which would have been achieved using an optimal allocation of search effort from the very beginning. This does not mean that time lost by non-optimal procedures at the beginning of the search can be recovered fully by subsequent search. It does mean that the effects of initial non-optimal allocation can be overcome as time progresses and the time penalties thereby minimized.

Estimates of the search effort required by various search systems to find *Scorpion* were made by determining the probability of detecting *Scorpion* as a function of time. Expected time to detection was also computed. Four alternative search systems were considered. These were one and two ships using camera only, and one and two ships using magnetometer and sonar simultaneously.

Figure 3 (submitted to the TAG August 1) shows probability of detection as a function of bottom time for one ship using magnetometer and sonar. Because of lack of information about the search area, expected time spent investigating false contacts was not included in this forecast. The extent of combined coverage is assumed to be 400 feet with probability of detection within that coverage of 0.5; track length is converted to time duration by assuming a speed of 1 knot. The cases of good and poor navigation are considered separately and the estimated expected time to detection is indicated.

It is interesting to note that the curve for poor navigation in Figure 3 (the case which corresponds closest to the actual search situation) indicates that the expected bottom time necessary to produce detection is between 35 and 45 days. The total length of track in the actual search including the research and development effort up to detection and excluding the identification effort around *M*8/3 is approximately 1026 miles or about 43 days of bottom time at the speed of 1 knot assumed in the computations.

Computer programs were developed to allow rapid and presumably more accurate computation of *SEP*, *LEP*, and a posteriori target location probabilities.

As mentioned above, throughout the search apparently man-made objects (artifacts) were photographed and careful records were made of their location. Estimates of the density of artifacts in various

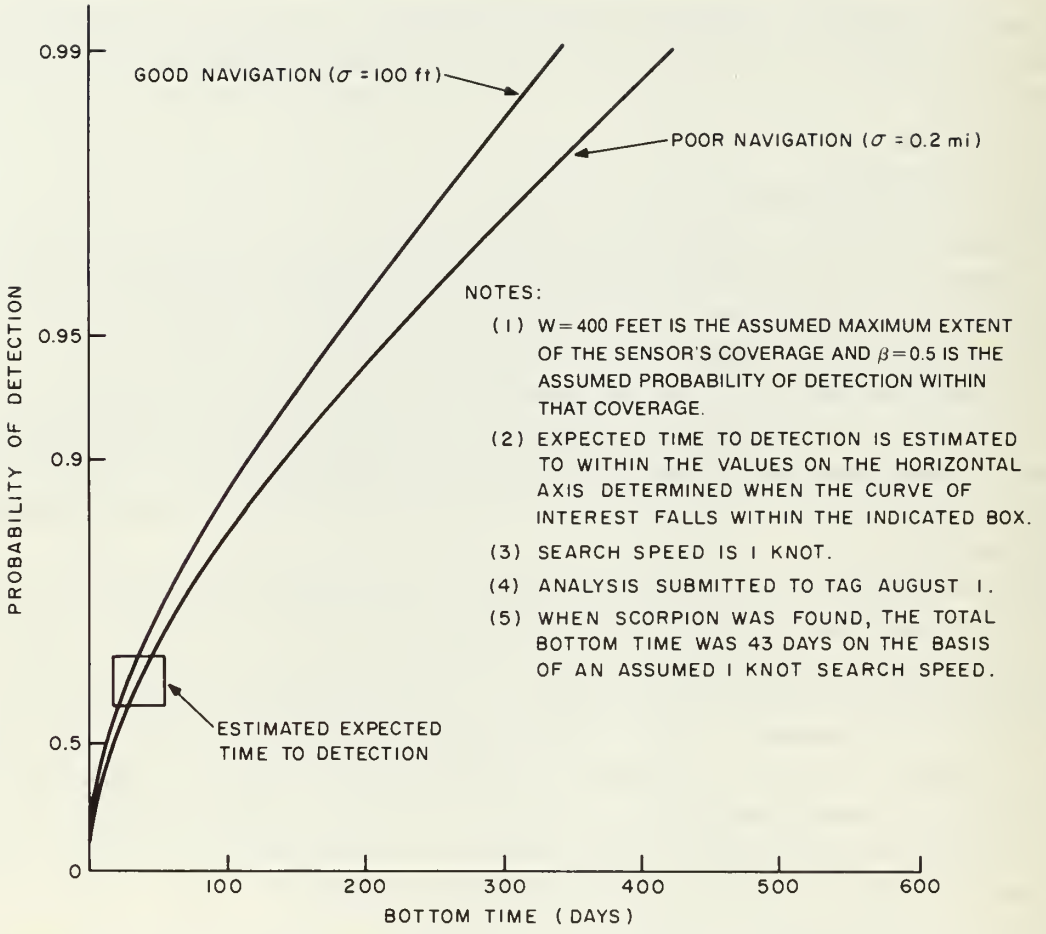

0.99
0.95
0.9
0.5
0
PROBABILITY OF DETECTION
GOOD NAVIGATION (σ = 100 ft)
POOR NAVIGATION (σ = 0.2 mi)
NOTES:
(1) W = 400 FEET IS THE ASSUMED MAXIMUM EXTENT OF THE SENSOR'S COVERAGE AND β = 0.5 IS THE ASSUMED PROBABILITY OF DETECTION WITHIN THAT COVERAGE.
(2) EXPECTED TIME TO DETECTION IS ESTIMATED TO WITHIN THE VALUES ON THE HORIZONTAL AXIS DETERMINED WHEN THE CURVE OF INTEREST FALLS WITHIN THE INDICATED BOX.
(3) SEARCH SPEED IS 1 KNOT.
(4) ANALYSIS SUBMITTED TO TAG AUGUST 1.
(5) WHEN SCORPION WAS FOUND, THE TOTAL BOTTOM TIME WAS 43 DAYS ON THE BASIS OF AN ASSUMED 1 KNOT SEARCH SPEED.
ESTIMATED EXPECTED TIME TO DETECTION
100
200
300
400
500
600
BOTTOM TIME (DAYS)

FIGURE 3. Probability of detection for one ship using magnetometer and sonar ($W = 400$ ft and $\beta = 0.5$)

rectangles were made for the purpose of determining whether a pattern was present which would require revision of the a priori distribution. These estimates and associated confidence bounds depended upon the value of visual detection probability in a cell as well as upon the number of artifacts actually observed.

In addition, the estimates were useful in determining the rectangles of highest artifact density; recommendations were presented for testing new camera techniques in areas most likely to contain objects of interest.

Because of the uncertainty in the capabilities of the nonvisual sensors and in particular, because of uncertainties in sweep width, analysis was performed to determine the penalties associated with incorrectly estimating sensor capabilities and to find optimal plans in the face of such uncertainty. It was shown that when no false contacts are present, finding an optimal plan for a sensor with uncertain sweep width (see Reference [2]) may be reduced to a standard separable allocation problem which may be attacked by the same types of methods used to solve this problem with deterministic sweep width, e.g., a Neyman-Pearson constrained extremal method (see Reference [5]). When false contacts are present, more general methods, such as those of [3] and [4], must be used to find optimal plans for search with uncertain sweep width and multiple sensors.

After-Search Estimates of Search Effectiveness

By examining the track charts for the runs during which *Scorpion* was detected and photographed, it is possible to make crude sensor sweep width estimates against the actual target of the search. With these estimates one may calculate, using hindsight, the effectiveness of the search effort as it developed in time. We emphasize that these estimates apply to *Scorpion* as it was actually found (i.e., partially buried). They may or may not be appropriate in another search.

The target of the search is taken as the largest piece of *Scorpion*. The sweep widths in Table 2 are used in our after-search assessment of effectiveness.

The after-search estimates of *SEP* for the search through runs 38, 63, 71, and 74 (*Scorpion* was photographed on run 75) are given in Table 3, based on the sweep widths of Table 2.

Scorpion's position is estimated to be in *F*6 and therefore the *LEP*'s in this rectangle are of particular interest. Table 3 presents these *LEP*'s throughout the search based on the after-search estimates of sensor sweep widths given in Table 2. It is interesting to compare these sweep widths with the estimates available to the on-scene analysts particularly for the non-visual sensor. The latter estimates were typically 800 feet and 400 feet for sonar and magnetometer, respectively.

TABLE 2. *After-Search Estimates of Sensor Sweep Widths*

	Largest-Piece Target		
	M[b]	β[b]	W[b]
Camera	250 ft	0.74	185 ft
Magnetometer	400 ft	0.55	220 ft
Sonar[a]	0	0	0

[a] The value 0 for the sonar sweep width against *Scorpion* is based on the failure of one of the two sonars to detect *Scorpion*. No data are available pertaining to the performance of the other sonar against the *Scorpion* as found.

TABLE 3. *After-Search Estimates of SEP and LEP in F6 Throughout the Search*

	Runs			
	1–38	1–63	1–71	1–74
	Termination Date			
	Aug. 6	Sept. 10	Oct. 9	Oct. 28
SEP	0.26	0.29	0.45	0.48
LEP in *F*6	0.46	0.47	0.65	0.74

[b] M indicates the maximum extent of sensor coverage and β indicates the probability of detection within the coverage; $W = \beta M$. The sweep width of the camera takes into account the length of the target.

Appraisal of Operations Analysis Effort

The a priori distribution is noted to be remarkably accurate with the location of *Scorpion* being placed 260 yards or less from the rectangle *E*5 which had the highest a priori probability of containing

Scorpion. Using the a priori distribution, estimates were made of the expected search effort required to find *Scorpion*. Although based on rather uncertain information, these estimates also proved to be quite accurate. Unfortunately, the a priori distribution was not available on scene until August 12.

As noted above, the a priori distribution combined with the *LEP*'s tended to give good advice as to where to search despite the high *LEP*'s resulting from the optimistic sweep width estimates.

The documentation and evaluation of search effort especially in the form of *LEP*'s and *SEP*'s provided a clear and concise method of describing the results of the search to the Technical Advisory Group.

Throughout the search there was no generally agreed-upon objective method of deciding when to investigate contacts nor was there any generally agreed upon criterion for defining a contact. This caused many possible contacts not to be recorded. It also appears in retrospect that one or more contacts in *F*6 should have been regarded as worthy of further investigation and treated as such in the *LEP* computation. This would have considerably reduced *LEP* in *F*6.

In view of the difficulties presented in the *Scorpion* search, lack of knowledge of the capabilities of the nonoptical sensors against various targets appears to be the major source of uncertainty in measuring search effectiveness and in optimal allocation of search effort. Because of the subsequent usefulness of the analysis of a few passes near the actual location of *Scorpion*, it is considered worth expending at least a modest amount of effort at the beginning of the search to make rough estimates of the sensor capabilities against objects which are similar to the target of the search.

The accuracy of the *SEP* computations was degraded to some extent by errors in the track charts introduced by uncertainty in the true locations of the navigational transponders. As the search progressed, the positions of these transponders became better known due to an increasing number of satellite cross checks. It was intended to revise the track charts during the planned winter operational pause and recompute the *LEP*'s and *SEP* accordingly. *Scorpion*, of course, was found before this effort could take place.

Improvement in communications and transmission of substantive information would help to improve search efficiency. Time lags of roughly three weeks to one month were experienced between the attainment of results stateside (e.g., the a priori distribution) and their application on scene.

One of the principal needs for improved analytic techniques arises in the area of contact investigation. In particular, it remains a problem to determine the proper allocation of effort between contact investigation and broad search in complicated searches involving many different types of sensors. Some progress on this problem has been made in [4].

In view of the uncertainties encountered regarding sensor capabilities in the *Scorpion* search and others (e.g., the H-bomb search), it is apparent that more consideration should be given to the problem of search optimization in the face of uncertainty in sensor sweep widths and other search parameters.

Appendix A

SEARCH EFFECTIVENESS PROBABILITY

This appendix discusses the computation of search effectiveness probability (*SEP*), the measure of effectiveness used during the *Scorpion* search. In contrast to the H-bomb search, where *SEP* was also used, performance of different sensors cannot be treated as statistically independent, since several sensors often were used from approximately the same physical location.

Let E_j, called local effectiveness probability (LEP), be the probability that if the target were in R_j (the j^{th} rectangle), it would have been found (detected and visually identified) by the search effort placed in R_j. (An exception to this definition occurs when no contacts have been generated and is discussed below.) The target might be found during the investigation of a contact generated by a nonvisual sensor or during the general visual search. Search effectiveness probability (SEP) is given by (N is the number of rectangles)

$$SEP = \sum_{j=1}^{N} L_j E_j,$$

where L_j is the a priori probability that the target is in the j^{th} rectangle.

In order to calculate the probabilities E_j, let D_j be the probability that if *Scorpion* were in rectangle j, it would have been detected either visually or nonvisually. Let P_j be the probability that if *Scorpion* were in rectangle j, it would have been detected (and it is assumed identified) by the general camera search (i.e., search not directed toward identifying a specific contact). Define C_j to be the probability that if the *Scorpion* were one of the nonvisual contacts in rectangle j, it would have been identified. If there are no nonvisual contacts in rectangle j, C_j is set equal to 1 by convention.

Then

$$E_j = P_j + C_j(D_j - P_j). \tag{A-1}$$

Let L_j' be defined by

$$L_j' = \frac{L_j(1-E_j)}{1-SEP}, \tag{A-2}$$

so that L_j' gives the a posteriori probability that *Scorpion* was in the j^{th} rectangle given the search up to the time at which the E_j's were computed. These are the numbers, considering the search to date, which are used to guide the search to the areas of highest probability of containing the target.

When $C_j = 1$ in (A–1), $E_j = D_j$ and LEP is the probability that at least one sensor (visual or nonvisual) detects, rather than the probability of detection and visual identification. This convention for C_j leads to the correct formula for the a posteriori probability distribution (A–2), but makes it difficult to give an intuitively appealing definition of SEP.

It can be shown that

$$\frac{SEP-Q}{1-Q} = Pr\left\{\begin{array}{l}\text{Target is detected and} \\ \text{visually identified}\end{array} \middle| \begin{array}{l}\text{Contacts at the locations} \\ \text{observed in the search}\end{array}\right\}, \tag{A-3}$$

where for $S_N = \{j\colon \text{no contacts in } R_j\}$

$$Q = \sum_{j \epsilon S_N} L_j Pr\left\{\begin{array}{l}\text{Target is detected in } R_j \text{ by} \\ \text{a } \textit{non-visual} \text{ sensor}\end{array} \middle| \text{ Target within } R_j\right\}.$$

The conditional probability of detection and identification given by (A–3) probably provides a better basis for terminating an unsuccessful search in the presence of false targets than does SEP. The derivation of (A–3) is omitted since it is rather lengthy and would take us beyond the expository nature of the present paper.

In order to compute D_j and P_j, track length records are used which give the amount of track of

each sensor group in each search rectangle. A sensor group is a combination of sensors which are used simultaneously. A list of sensor groups used in the *Scorpion* search is given in Table A–1.

TABLE A–1. *Definition of Sensor Groups for Scorpion Search*

Sensor Group Designation [a] [b]	Description
Single Sensor 1	Camera
Single Sensor 2	Magnetometer
Single Sensor 3	First Sonar (Short Range–SR)
Single Sensor 4	First Sonar (Long Range–LR)
Single Sensor 5	Second Sonar
Single Sensor 6	.....
7	Camera and Magnetometer
8	Camera and First Sonar (SR)
9	.....
10	Camera and Second Sonar
11	.....
12	Magnetometer and First Sonar (SR)
13	Magnetometer and First Sonar (LR)
14	Magnetometer and Second Sonar
15	.....
16	Camera, Magnetometer, and First Sonar (SR)
17	.....
18	Camera, Magnetometer, and Second Sonar

[a] Gaps in the table correspond to provision made for the introduction of new sensor groups.
[b] Only the sensors actually mentioned belong in the group.

Let Ω_i be the sweep width of the i^{th} sensor group and let Λ_{ij} be the amount of track length of the i^{th} sensor group in the j^{th} rectangle. Because of the large navigational uncertainties and lack of regular search plans, the probability $D_j^{(i)}$ of the i^{th} sensor group detecting the *Scorpion* given that it is the j^{th} rectangle is computed by the random search formula of Koopman [1]. That is,

$$D_j^{(i)} = 1 - \exp\left(\frac{-\Lambda_{ij}\Omega_j}{A_j}\right),$$

where A_j is the area of the j^{th} rectangle ($A_j = 0.84$ square miles for the *Scorpion* search). The above formula assumes the effort is uniformly, but randomly, distributed in the area. To compute D_j, the probability that at least one of the sensor groups would detect *Scorpion* given that it is in the j^{th} rectangle, we assume that the sensor *groups* are mutually statistically independent. Thus

$$D_j = 1 - \prod_{i=1}^{18} (1 - D_j^{(i)}).$$

(Note there are 18 sensor groups in Table A–1 not all of which were actually used in the search.)

The probabilities P_j are calculated as follows: Let $\Lambda_j^{(c)}$ be the amount of camera track length in the j^{th} grid rectangle and let W_c be the camera sweep width. Then P_j is given by

$$P_j = 1 - \exp\left(\frac{-\Lambda_j^{(c)} W_c}{A_j}\right).$$

The track length $\Lambda_j^{(c)}$ is computed by the equation

$$\Lambda_j^{(c)}=\Lambda_{1j}+\Lambda_{7j}+\Lambda_{8j}+\Lambda_{10j}+\Lambda_{16j}+\Lambda_{18j}.$$

Let K_j be the number of contacts in the j^{th} rectangle. For $K_j \neq 0$, let γ_{kj} be the conditional probability that the target would have been identified by contact investigation, given that the target is the k^{th} contact in the j^{th} rectangle (see the computation of γ_{kj} in the case of contact *M*8/3 discussed in appendix B).

The contact investigation probabilities C_j are computed by

$$C_j=\begin{cases}\dfrac{1}{K_j}\displaystyle\sum_{k=1}^{K_j}\gamma_{kj} & \text{for } K_j\neq 0\\ 1 & \text{otherwise.}\end{cases}$$

Appendix B

INVESTIGATION OF CONTACT M8/3

This appendix discusses operations analysis connected with the investigation of the contact designated *M*8/3. This was the only contact other than the final contact actually produced by the *Scorpion* which was systematically investigated. Operations analysis provided a basis for allocating contact investigation effort and contributed to the decision to abandon the investigation.

A Priori Contact Location Distribution

In order to systematically investigate *M*8/3 a circular normal a priori contact location distribution with standard deviation $\sigma=1200$ ft in any direction was postulated. The value chosen for σ was based on the navigational uncertainty at the time the contact was made and on the uncertainty in returning to the charted location (i.e., posit) of *M*8/3.

Upon returning to the search area on August 14, coded transponders were planted in the vicinity of the posit of *M*8/3. The locations of these transponders were calculated by use of several satellite fixes. All navigation about *M*8/3 was then performed with respect to these transponders, and a check of the navigational errors for this system indicated that the navigational uncertainty had a standard deviation σ_N of about 300 ft in any direction.

Because of the improved navigational accuracy, a microgrid was constructed with center at the posit of *M*8/3. The grid was composed of 64 cells, 1000 ft on a side. The cells were indexed by the ordered pairs (i, j) $i=1, \ldots, 8$ and $j=1, \ldots, 8$. Using a circular normal distribution with $\sigma=1200$ ft for the location of *M*8/3, L_{ij}^{μ}, the probability that *M*8/3 was contained in the $(i, j)^{th}$ grid cell, was computed and used to approximate the circular normal distribution above by assuming that the location of *M*8/3 was contained in the $(i, j)^{th}$ grid cell with probability L_{ij}^{μ}.

Allocation of Magnetometer Effort

The magnetometer local detection probability $\tilde{D}_{ij}^{m}$ in the $(i, j)^{th}$ cell is the probability that the magnetometer would have detected *M*8/3 given that *M*8/3 was in that cell. The magnetometer was assumed to have a definite range law; that is, it would detect with probability $\beta=1$ if *M*8/3 were at lateral ranges of 200 feet or less and would not detect beyond that. This gave the magnetometer a sweep width W_m of 400 feet. The search within each cell was evaluated as random. That is, if Λ_{ij}^{μ} was the

amount of track length placed in the $(i,j)^{th}$ cell and A_{ij}^{μ} the area of that cell, then

$$\tilde{D}_{ij}^{m} = 1 - \exp\left[-\left(\frac{W_m \Lambda_{ij}^{\mu}}{A_{ij}^{\mu}}\right)\right]. \tag{B-1}$$

Detection effectiveness probability (*DEP*) is the probability that if *M*8/3 were the *Scorpion*, it would have been detected by the magnetometer during the microsearch. That is,

$$DEP = \sum_{i=1}^{8} \sum_{j=1}^{8} L_{ij}^{\mu} \tilde{D}_{ij}^{m}. \tag{B-2}$$

Every two to three runs *DEP* was computed and posteriori contact location probabilities were calculated by

$$\tilde{L}_{ij}^{\mu} = \frac{L_{ij}^{\mu}\,(1-\tilde{D}_{ij}^{m})}{DEP}.$$

The magnetometer search effort was directed to the cells having the highest posterior probabilities. This provided a rough optimization of magnetometer detection effort.

The *DEP* resulting from runs 39–51 is 0.95. Run 52 which was cut short because of weak batteries in the bottom transponders near *M*8/3 is not included in the calculations of effectiveness.

Identification of Contacts Found During Runs 39–51

During the search about *M*8/3, all significant magnetometer contacts were found within a circle of radius 650 feet centered at *PIX* 42. *PIX* 42 is located approximately 1300 ft south and east of the posit of *M*8/3 and is so called because it is the location of a group of rocks photographed during run 42 at the same time a magnetic anomaly was detected.

An intensive camera search was performed inside this circle. The circle was divided into quadrants, and camera detection probabilities were calculated within each of these quadrants. In order to calculate these probabilities, it was assumed that the camera had a sweep width $W_c = 75$ ft and that detection and identification occurs with probability 1 across this sweep width (i.e., $\beta = 1$). The value of W_c that was used is an average sweep width derived from data available on scene. The effectiveness of the camera search was evaluated as random in each quadrant, and the appropriate analog of (*B*–1) was used to calculate the camera detection probabilities in each quadrant.

The camera detection probabilities for the four quadrants were averaged to compute $C^{\mu} = 0.91$ the probability that if one of the micro-contacts were the *Scorpion*, it would have been identified.

Thus if *M*8/3 were *Scorpion*,

$$C_j = Pr\{\textit{Scorpion} \text{ would have been identified} \mid \textit{Scorpion} \text{ among micro contacts}\}.$$

$$Pr\{\textit{Scorpion} \text{ among micro contacts}\}$$

$$= C^{\mu}\, DEP = 0.91 \times 0.95 = 0.86.$$

It is this value of 0.86 which is used for C_j in *D*8 (the rectangle containing *M*8/3). This high value for C_j contributed to the commodore's decision to abandon the investigation of *M*8/3 after run 52.

ACKNOWLEDGMENTS

The authors wish to make several acknowledgments. Particular gratitude is expressed for the scientific leadership of Dr. John P. Craven, Director of the stateside Technical Advisory Group formed for the search; the general view is shared that his insight and perseverance kept the search going, and he contributed a vital (and highly successful) part of the operations analysis. It was a pleasure to work under the command of the search commodores—their approaches were varied, but in general it was their interest in the usefulness of search analysis that made the operations research efforts worthwhile. The splendid cooperation in logistical support, etc., received from the staff of Commander, Submarine Forces, Atlantic Fleet, under whose command this successful operation was entirely carried out, is very much appreciated. Submarine Development Group One officers did a very creditable job of rapid mastery of the relevant operations analysis techniques during stateside briefings and of carrying on the calculations on scene in the latter part of the search. Deep respect is expressed for the splendid ocean engineering work done by the Naval Research Laboratory (NRL) scientific teams working aboard *Mizar*—the feat of towing and locating the sensor sled close off the ocean bottom in some 10,000 feet of water for weeks on end over the five months search period, with rather low equipment downtime, reflects great skill and dedication. Moreover, their Chief Scientist, Mr. C. L. Buchanan, was directing the search during the research and development phase during which the *Scorpion* was finally found. Finally, it was a pleasure to cooperate with Dr. Frank A. Andrews of Catholic University who assisted the Technical Advisory Group as an Ocean Systems, Inc., consultant; having commanded the *Thresher* search as a Captain, USN, and having consulted to the H-bomb search, he has an unusually comprehensive knowledge of and experience in the operational engineering, and analytical aspects of deep submerged search.

REFERENCES

[1] Koopman, B. O., "The Theory of Search: III Optimum Distribution of Searching Effort," Operations Research **5**, 613–626 (1957).

[2] Richardson, H. R. and B. Belkin, "Optimal Search With Uncertain Sweep Width," submitted for publication.

[3] Stone, L. D. and J. A. Stanshine, "Optimal Search Using Uninterrupted Contact Investigation," SIAM Journal on Applied Mathematics, **20**, 241–263 (1971).

[4] Stone, L. D., J. A. Stanshine, and C. A. Persinger, "Optimal Search With Poisson Distributed False Targets," submitted for publication.

[5] Wagner, D. H., "Nonlinear Functional Versions of the Neyman-Pearson Lemma," SIAM Review **11**, 52–65 (1969).

OPTIMUM SEQUENTIAL SEARCH WITH DISCRETE LOCATIONS AND RANDOM ACCEPTANCE ERRORS

Mark W. Smith

Texas Instruments Incorporated

and

John E. Walsh

Southern Methodist University

ABSTRACT

Much work has been done in search theory; however, very little effort has occurred where an object's presence at a location can be accepted when no object is present there. The case analyzed is of this type. The number of locations is finite, a single object is stationary at one location, and only one location is observed each step of the search. The object's location has a known prior probability distribution. Also known are the conditional probability of acceptance given the object's absence (small) and the conditional probability of rejection given the object's presence (not too large); these probabilities remain fixed for all searching and locations. The class of sequential search policies which terminate the search at the first acceptance is assumed. A single two-part optimization criterion is considered. The search sequence is found which (i) minimizes the probability of obtaining n rejections in the first n steps for all n, and (ii) maximizes the probability that the first acceptance occurs within the first n steps and occurs at the object's location for all n. The optimum sequential search policy specifies that the next location observed is one with the largest posterior probability of the object's presence (evaluated after each step from Bayes Rule) and that the object is at the first location where acceptance occurs. Placement at the first acceptance seems appropriate when the conditional probability of acceptance given the object's absence is sufficiently small. Search always terminates (with probability one). Optimum truncated sequential policies are also considered. Methods are given for evaluating some pertinent properties and for investigating the possibility that no object occurs at any location.

INTRODUCTION AND DISCUSSION

Development of efficient search policies has received much attention [1], [2]. Virtually without exception, however, these results are for situations where observation (for detection of an object) is such that there is zero probability of accepting an object's presence at a place where no object is located. In practice, many important cases occur where this probability is nonzero (although it often may be small).

This paper considers a case where the conditional probabilities of acceptance (rejection), given the presence (absence) of an object, are neither zero nor unity. The search space is discrete, with a finite number N of possible locations. There is only one object and it remains stationary at one of the locations. Given is a valid prior probability distribution for the true location of the object. Also, the conditional probabilities of acceptance and rejection have known values that remain the same for all locations and all steps of the search (perhaps adjustable prior to the search).

The problem is to determine an optimum search policy for identifying an object location subject to observing one location at each step of the search (the same location could be observed at successive

steps). Only search policies which terminate the search at the first acceptance are considered. A single two-part optimization criterion is considered. A search sequence is found which

(i) minimizes the probability of obtaining n rejections in the first n steps for all n, and

(ii) maximizes the probability that the first acceptance occurs within the first n steps and occurs at the object's location for all n.

The principal interest is in situations where the conditional probability of acceptance given the object is absent (denoted by α) is small. Then, if the conditional probability of rejection given that the object is present (denoted by β) is not too large, there is a good chance that the first location accepted is the true location. Such situations are found to occur when $(1-\beta)/(N-1)\alpha$ is much larger than unity. A restricted class of policies which identifies the first location accepted as that of the object seems appropriate under these circumstances (and corresponds to previous results for the limiting situation of $\alpha=0$). The properties obtained for these policies actually hold when $\alpha,\beta < 1/2$, but may not be very desirable unless $(1-\beta)/(N-1)\alpha$ is sufficiently large.

Consider the details of the optimum policy. At the first step, a location with largest value for the prior probability of the object's presence is observed. If the location is accepted, the object is said to be there and the search is completed. When rejection occurs, the posterior probabilities of the object's location are computed using Bayes Rule (from the location observed, occurrence of rejection, the prior distribution, α, and β). The next location observed is one with the largest of these posterior probabilities. This procedure is continued until an acceptance, which ultimately happens (with probability one). Then search is terminated and the identified location is that where acceptance occurred. When more than one location has the largest probability value, one of them is randomly chosen on a equally-likely basis.

The probability of correctly selecting the object's location can be made as near unity as desired if α and/or β can be modified prior to beginning the search. For fixed β, it is found that this can be accomplished by having α sufficiently small; however, the largest usable value for α increases as β decreases, so that having β small is desirable. Sometimes α and β are related so that decreasing the value of either one increases the value of the other. Then a "best" combination of values can often be determined for α and β. A specific example involving communications search is given in Ref. [3]. Also, some potential fields of practical application for the optimum sequential search policies, with emphasis on electronic problems, are mentioned in Ref. [3]. The material of Ref. [3] (unpublished) provides the basis for much of this paper and will be referred to again when derivations are considered.

Optimum truncated sequential policies are also obtained. Here, a maximum is specified for the number of steps. The search policy is the same as that already given through this number of steps. When rejection occurs for all the steps, search ends and the identified location is one with largest posterior probability (computed after the last step). The probability of correct decision is less than for the policy without truncation but has very nearly the same value when the maximum for the number of steps is sufficiently large. Computations can be made to determine a suitable maximum number of steps.

Exact expressions are developed for some properties such as the probability of correct location, the expected number of steps to search termination, and the probability of search termination within a stated number of steps. These expressions can be evaluated as closely as desired (exactly when the number of steps is limited) by direct utilization of the optimum policy being considered. Upper and lower bounds are developed for the properties considered that involve an unlimited number of steps. These

bounds tend to be the same value with an increasing number of steps and are useful in deciding when an approximation to an evaluation is sufficiently accurate.

Part of the approach to providing bounds consists in using the uniform distribution (probability 1/N for each location) as the prior distribution. That is, part of one of the two bounds is evaluated on the assumption of a uniform prior distribution. In particular, a probability of correct decision evaluated for the situation of a uniform prior distribution is a lower bound for values of this probability. Also, an expected number of steps to search termination obtained on this basis provides an upper bound, etc. A desirable feature of bounds determined from the uniform distribution is that they often have reasonably uncomplicated expressions of a closed form. Then, sufficient conditions on α and β for obtaining desired values of properties can be determined from examination of an uncomplicated function of α and β.

Sometimes an object is believed to be at some one of the locations but its occurrence is not certain. That is, the possibility exists that no object is present at any location. Then a form of truncated policy should be used, since the policy with an unlimited number of steps is certain to ultimately identify a location even when no object occurs. For such a truncated policy, the conclusion that there is no object is adopted after the maximum number of steps is reached without an acceptance. Acceptance at any of the steps implies that there is an object and identifies its location (search also stops). Here the prior distribution used for a search policy is conditional (it is given that the object occurs). Ideally, the policy used should provide a good chance of claiming that no object occurs when this is the case and also of finding it (including location) when there is an object. Such an ideal situation is found to occur when α is sufficiently small and the maximum for the number of steps is suitably chosen.

Statement of notation and results concerning the properties considered occurs in the next section. The final section is concerned with the situation where the possibility exists that no object is present at any location.

NOTATION AND RESULTS

The locations are identified by $l (l = 1, \ldots, N)$. The location numbering is such that

$$p_0(1) \geq p_0(2) \geq \ldots \geq p_0(N),$$

where $p_0(l)$, nonzero, is the prior probability that the object is at location l and

$$p_0(1) + \ldots + p_0(N) = 1.$$

The steps of the search are $i = 1, 2, \ldots$ and one location is observed, for possible detection of whether the object is there, at each step. $m_i(l)$ equals the number of times that location l has been observed during steps $1, \ldots, i-1$. The location e_i is such that, for use of an optimal policy,

$$K_i p_0(e_i) \equiv [\beta/(1-\alpha)]^{m_i(e_i)} p_0(e_i) = \max_l [\beta/(1-\alpha)]^{m_i(l)} p_0(l)$$

and is the location observed at the i^{th} step of the optimum policy. If the maximum occurs for more than one location, the value of e_i is chosen randomly so that all possibilities are equally likely (to avoid bias). For evaluation of properties, however, the value used for e_i is the smallest number among the possibilities.

Knowledge of α, β, N, and the $p_o(l)$ combined with methods of evaluating e_i and K_i over all i is sufficient both for application of an optimum policy and for evaluating properties of this policy. For a truncated optimum policy, the maximum number of steps M is also given.

Now consider statements and verification of properties. In all cases, optimum policies are used. Detailed derivations of the various expressions used in verification for nearly all of the results have been given in Ref. [3]. Thus, nearly all of the verification given here is only sketched, and often uses needed expressions without consideration of their derivations.

THEOREM 1. *The minimum probability that there are rejections at all of the first* n *steps is* $1-P'_n$, *where*

$$P'_n = 1-(1-\alpha)^n + (1-\alpha-\beta)(1-\alpha)^{n-1}\sum_{i=1}^{n} K_i p_0(e_i)$$

is the probability that search will terminate on or before the n^{th} *step. Use of an optimum policy minimizes* $1-P'_n$.

Verification. Mathematical induction is used to establish the form of the probability that rejections occur at the first n steps for any n. For the first step ($n=1$) and location l_1 observed, the probability of rejection is easily seen to be

$$(1-\alpha-\beta)[(1-\alpha)/(1-\alpha-\beta)-p_0(l_1)].$$

Now consider $n \geq 2$ and assume that the probability of having rejections at the first $n-1$ steps, with location l_i observed for the i^{th} step, is

$$(1-\alpha-\beta)(1-\alpha)^{n-2}\left[\frac{1-\alpha}{1-\alpha-\beta}-\sum_{i=1}^{n-1}\left(\frac{\beta}{1-\alpha}\right)^{m_i(l_i)}p_0(l_i)\right]. \tag{1}$$

The induction is completed by showing that (1) implies the expression

$$(1-\alpha-\beta)(1-\alpha)^{n-1}\left[\frac{1-\alpha}{1-\alpha-\beta}-\sum_{i=1}^{n}\left(\frac{\beta}{1-\alpha}\right)^{m_i(l_i)}p_0(l_i)\right] \tag{2}$$

for the probability of rejection on the first n steps, where location l_n is observed at the n^{th} step.

For verification of (2) from (1), the probability of rejection on steps $1, \ldots, n$ is expressed as the probability of rejection on the first $n-1$ steps multiplied by the conditional probability of rejection on the n^{th} step given that the first $n-1$ steps yielded rejections. The conditional probability is found to be

$$(1-\alpha)\left[\frac{1-\alpha}{1-\alpha-\beta}-\sum_{i=1}^{n}\left(\frac{\beta}{1-\alpha}\right)^{m_i(l_i)}p_0(l_i)\right]\left[\frac{1-\alpha}{1-\alpha-\beta}-\sum_{i=1}^{n-1}\left(\frac{\beta}{1-\alpha}\right)^{m_i(l_i)}p_0(l_i)\right]^{-1}.$$

Multiplication of this probability by the assumed expression (1) yields (2).

General verification of (2) implies that, for any number of steps, an optimum policy minimizes the probability of all rejections. That is, (2) is minimum for $l_i = e_i$, $(i=1, \ldots, n)$ since

$$K_i p_0(e_i) \equiv \max_{l_i}\left[\beta/(1-\alpha)\right]^{m_i(l_i)} p_0(l_i)$$

for all possible choices of l_i and for all i.

Since the last two terms of P'_n tend to zero as $n \to \infty$, there is unit probability that search terminates. In fact, search terminates with probability one for any search policy with the property that, ultimately, every location is observed an unlimited number of times.

THEOREM 2. *For any number* n *of steps, an optimum policy maximizes the probability that the first acceptance occurs within the first* n *steps and is at the object's location. The value of this maximum probability is*

$$P_n = (1-\beta)\sum_{i=1}^{n}(1-\alpha)^{i-1}K_i p_0(e_i).$$

Thus, for an unlimited number of steps, the maximum probability that the first acceptance is at the object location is

$$P_\infty = (1-\beta)\sum_{i=1}^{\infty}(1-\alpha)^{i-1}K_i p_0(e_i)$$

and occurs when an optimum policy is used.

Outline of Verification. Suppose that location l_i is observed at the i^{th} step ($i=1, \ldots, n$). The probability that the first acceptance occurs at step i and is at the object's location equals the probability of rejection on the first $i-1$ steps multiplied by two conditional probabilities. One is the conditional probability of acceptance at step i given that the object is at l_i and that steps $1, \ldots, i-1$ yielded rejections. This probability equals $1-\beta$. The other is the conditional probability that the object is at l_i given that steps $1, \ldots, i-1$ yielded rejections. This probability is found to be

$$\left(\frac{1-\alpha}{1-\alpha-\beta}\right)\left(\frac{\beta}{1-\alpha}\right)^{m_i(l_i)}p_0(l_i)\left[\frac{1-\alpha}{1-\alpha-\beta}-\sum_{j=1}^{i-1}\left(\frac{\beta}{1-\alpha}\right)^{m_j(l_j)}p_0(l_j)\right]^{-1}.$$

The product of the three probabilities simplifies to

$$(1-\beta)(1-\alpha)^{i-1}[\beta/(1-\alpha)]^{m_i(l_i)}p_0(l_i),$$

so that

$$(1-\beta)\sum_{i=1}^{n}(1-\alpha)^{i-1}[\beta/(1-\alpha)]^{m_i(l_i)}p_0(l_i)$$

is a general expression for the probability that the first acceptance occurs within the first n steps and is at the object's location.

Use of an optimum policy maximizes (3) for every value of n, since use of $l_i = e_i$ maximizes $[\beta/(1-\alpha)]^{m_i(l_i)}p_0(l_i)$ for all i. When the number of steps is unlimited, this result shows that an optimum policy results in the maximum probability that the first acceptance occurs at the object's location.

THEOREM 3. *The minimum of the expected number of steps to termination of search, when the number of steps is unlimited, equals*

$$\underline{L} = \frac{1}{\alpha}\left[1-(1-\alpha-\beta)\sum_{i=1}^{\infty}(1-\alpha)^{i-1}K_i p_0(e_i)\right]$$

and occurs when an optimum policy is used.

Verification. A general expression for the expected number of steps to the first acceptance, with location l_i observed at step i, is

$$\frac{1}{\alpha}\left\{1-(1-\alpha-\beta)\sum_{i=1}^{\infty}(1-\alpha)^{i-1}[\beta/(1-\alpha)]^{m_i(l_i)}p_0(l_i)\right\}.$$

Use of an optimum policy minimizes this expression and yields $\underline{L}$. For a perfect detection device ($\alpha=\beta=0$), the expected number of steps to search termination is

$$\sum_{l=1}^{N} lp_0(l).$$

The median number of steps to search termination is (conservatively) the smallest n such that $P'_n \geq 1/2$.

Now, consider the general situation with truncation. For $n \leq M$, the probability that the first acceptance occurs on or before the nth step is P_n, as given for no truncation.

THEOREM 4. *For truncation at* M *steps, the maximum probability of correctly locating the object is*

$$P_M + (1-\alpha)^M K_{M+1} p_0(e_{M+1}), \tag{4}$$

and occurs for use of an optimum policy.

Verification. For truncation at M steps, the probability of a correct decision is

$$(1-\alpha)^M[\beta/(1-\alpha)]^{m_{M+1}(l_{M+1})}p_0(l_{M+1}) + (1-\beta)\sum_{i=1}^{M}(1-\alpha)^{i-1}[\beta/(1-\alpha)]^{m_i(l_i)}p_0(l_i)$$

for the general case. Use of an optimum policy, so that $l_i = e_i$, $(i=1, \ldots, M+1)$, evidently maximizes this probability and yields (4).

For $n < M$, the probability that search will end on or before step n is P'_n. The expected number of steps to termination is

$$M - \sum_{i=1}^{M-1} P'_i.$$

The median number of steps to termination can (conservatively) be determined as the smallest n satisfying $P'_n \geq 1/2$, and equals M if no such n exists.

Next, consider the special situation of a uniform distribution and use of an optimum policy. Examination of the results for the general situation shows that these are determined when expressions of the forms $K_{rN+s}p_0(e_{rN+s})$,

$$\sum_{i=1}^{rN+s}(1-\alpha)^i K_i p_0(e_i), \text{ and } \sum_{i=1}^{rN+s} K_i p_0(e_i)$$

are evaluated for the uniform distribution, where $0 \leq r \leq \infty$, $1 \leq s < N$. It is found, directly from the fact that the locations are observed in the cyclic order $1, 2, \ldots, N, 1, 2, \ldots, N, 1, 2, \ldots$, that

$$K_{rN+s}p_0(e_{rN+s}) = \frac{1}{N}[\beta/(1-\alpha)]^r,$$

$$\sum_{i=1}^{rN+s}(1-\alpha)^{i-1}K_i p_0(e_i) \equiv U_1(rN+s) = \frac{1}{N\alpha}\left[[1-(1-\alpha)^N][1-\beta^r(1-\alpha)^{r(N-1)}]/[1-\beta)\,(1-\alpha)^{N-1}\right.$$

$$\left. + \beta^r(1-\alpha)^{r(N-1)}[1-(1-\alpha)^s]\right],$$

$$\sum_{i=1}^{rN+s} K_i p_0(e_i) \equiv U_2(rN+s) = \{1-[\beta/(1-\alpha)]^r\}/[1-\beta/(1-\alpha)] + (s/N)[\beta/(1-\alpha)]^r,$$

and closed-form expressions for P_∞, P_n, P'_n, etc. can be directly stated in terms of α, β, N, and perhaps n or M.

Consideration of the uniform distribution is useful because of some inequalities that occur.

THEOREM 5. *The inequalities*

$$\sum_{i=r_1N+1}^{r_2N+s}(1-\alpha)^{i-1}K_ip_0(e_i) \geq U_1(r_2N+s)-U_1(r_1N),$$

$$\sum_{i=r_1N+1}^{r_2N+s}K_ip_0(e_i) \geq U_2(r_2N+s)-U_2(r_1N), \tag{5}$$

where, $0 \leq r_1 \leq r_2 \leq \infty$, $1 \leq s \leq N$, *are satisfied for all possible sets of values for the* $p_0(1)$. *Thus,*

$$\sum_{i=1}^{r_2N+s}(1-\alpha)^{i-1}K_ip_0(e_i) \geq \sum_{i=1}^{r_1N}(1-\alpha)^{i-1}K_ip_0(e_i)+U_1(r_2N+s)-U_1(r_1N)$$

and

$$\sum_{i=1}^{r_2N+s}K_ip_0(e_i) \geq \sum_{i=1}^{r_1N}K_ip_0(e_i)+U_2(r_2N+s)-U_2(r_1N).$$

Verification. The cyclic order of the search of locations is used in combination with the inequalities

$$K_ip_0(e_i) \geq K_{i+1}p_0(e_{i+1}) \qquad \text{and} \qquad K_{rN+s}p_0(e_{rN+s}) \geq [\beta/(1-\alpha)]^r p_0(s).$$

The quantity on the right side of one of the inequalities is at most equal to a value in a set over which the quantity on the left side is the maximum value. This, combined with the properties of the $p_0(l)$, shows that

$$\sum_{i=r_1N+1}^{r_2N}K_ip_0(e_i) \geq \sum_{j=r_1}^{r_2}[\beta/(1-\alpha)]^j = U_2(r_2N)-U_2(r_1N)$$

and that

$$\sum_{i=r_2N+1}^{r_2N+s}K_ip_0(e_i) \geq \frac{s}{N}[\beta/(1-\alpha)]^{r_2}.$$

In a similar fashion,

$$\sum_{i=r_1N+1}^{r_2N+s}(1-\alpha)^iK_ip_0(e_i) \geq U_1(r_2N+s)-U_2(r_1N),$$

is found to hold.

Suppose that r_2N+s is exceedingly large so that evaluation of $K_ip_0(e_i)$ would require much effort for that many steps. Often, a value occurs for r_1 such that r_1N is not excessive but large enough for the stated lower bound to be near the true value. The lower bound involving $U_1(r_2N+s)$ is especially useful for r_2 infinite.

Upper bounds are also needed if sufficiently accurate determination of approximate values is to occur. The upper bounds developed are given by

THEOREM 6.

$$\sum_{i=1}^{r_2N+s}(1-\alpha)^{i-1}K_ip_0(e_i) \leq \sum_{i=1}^{r_1N}(1-\alpha)^{i-1}K_ip_0(e_i)+(1/\alpha)K_{r_1N}p_0(e_{r_1N})[(1-\alpha)^{r_1N}-(1-\alpha)^{r_2N+s}]$$

and

$$\sum_{i=1}^{r_2N+s}K_ip_0(e_i) \leq \sum_{i=1}^{r_1N}K_ip_0(e_i)+[(r_2-r_1)N+sK_{r_1N}p_0(e_{r_1N})].$$

Verification. The inequality

$$K_i p_0(e_i) \geq K_j p_0(e_j), \qquad (i \leq j).$$

directly yields these upper bounds.

The upper bounds can be used, in combination with the lower bounds, to determine when $r_1 N$ is large enough to assure that a good approximation is obtained. The approximation could, for example, be the arithmetic average of the upper and lower bounds. The upper bounds are especially useful when

$$\sum_{i=1}^{r_2 N+s} (1-\alpha)^{i-1} K_i p_0(e_i)$$

is to be evaluated (maybe with r_2 infinite). The forms of upper bounds are also usable when $r_1 N$ is replaced by an integer in the range 1 to $r_2 N + s$.

These results for the two types of summations are directly usable in approximate evaluation of P_∞, P_n, P'_n, and $\underline{L}$. The exact results for the uniform distribution furnish lower bounds on the possible values of P_∞, P_n, and P'_n. They provide upper bounds on the possible values of $\underline{L}$, $M - \sum_{i=1}^{n_{M-1}} P'_i$, and the median number of steps to termination. These closed form expressions are reasonably uncomplicated and can be used to establish conditions on α and β that are sufficient to assure desired values of properties (at least a stated value for P_∞, P_n, or P'_n; at most a stated value for the expected or median number of steps to termination).

As an example, consider that at least a value P is desired for P_∞. Use of P_∞ for the uniform distribution provides the relation

$$\frac{1-\beta}{N\alpha} [1-(1-\alpha)^N]/[1-\beta(1-\alpha)^{N-1}] \geq P$$

For $N\alpha \ll 1$ and $\beta < 1$, this becomes, approximately,

$$(1-\beta)/(N-1)\alpha \geq (1+\beta)/2(1-P),$$

so that $(1-\beta)/(N-1)\alpha$ should be much greater than unity for P large. A somewhat related problem, not involving P_∞, would be to choose α and β so that the posterior probability (of correctness) for the accepted location is at least P irrespective of the step at which acceptance occurs. The minimizing posterior distribution at the step just before acceptance is the uniform distribution. Then the posterior probability of correct location after acceptance on the next step is seen to be

$$\alpha(1-\beta)[(1-\alpha-\beta)N\alpha(\alpha/(1-\alpha-\beta)+1/N)]^{-1},$$

since the situation is the same as that of a uniform prior distribution and acceptance at the first step. The requirement that this value is at least P provides the (equivalent) inequality

$$(1-\beta)/(N-1)\alpha \geq P/(1-P).$$

Here too, $(1-\beta)/(N-1)\alpha$ should be much greater than unity for P large.

UNCERTAIN PRESENCE OF ANY OBJECT

The policies considered here have a maximum of m steps. When an acceptance occurs on or before the m^{th} step, the presence of an object is declared at the location of the acceptance and search

stops. If rejections occur for all of the first m steps, the decision is that no object is present at any location.

The probability of deciding that no object is present equals $(1-\alpha)^m$ when this is the situation. That is, the locations are equivalent, so that the same probability is obtained for all search policies. The probability of a correct decision for this situation is large when α and m are such that $(1-\alpha)^m$ is large. For fixed m, this occurs as α decreases. For fixed α, the value of $(1-\alpha)^m$ increases as m decreases but never exceeds $1-\alpha$.

When an object is present, the probability of deciding an object occurs and also correctly locating it equals P_m. The value of P_m increases as m increases and, for fixed m, increases as α decreases. However, P_m also depends on β and decreases as β increases. For m large and $\beta < 1$, a sufficiently small α will make P_m as near unity as desired. However, for P_m to be at least a stated value, the largest allowable α increases as β decreases, so that having β small is desirable.

Suppose that β is given and that a combination of α and m such that $(1-\alpha)^m \geq p_1$ and $p_m \geq p_2$ is desired, where p_1 and p_2 are specified values. This can be determined on an iterative basis by selecting a small value for α and trying to choose m so that both inequalities hold. If more than one value of m accomplishes this, increase α and try again. If no value of m suffices, decrease α and try again. The iteration stops when an α value is reached such that the inequalities are satisfied for exactly one value of m.

ACKNOWLEDGMENT

Thanks are due to Texas Instruments Incorporated for support of Mark W. Smith, and to ONR Contract N00014–68–A–515 and Mobil Research and Development Corporation for partial support of John E. Walsh.

REFERENCES

[1] Dobbie, J. M., "A Survey of Search Literature," Operations Research *14*, 525–537 (1968).

[2] Enslow, P. H., Jr., "A Bibliography of Search Theory and Reconnaissance Theory Literature," Nav. Res. Log. Quart. *13*, 177–202 (1966).

[3] Smith, M. W., "An Optimum Discrete Space Sequential Search Procedure Which Considers False Alarm and False Dismissal Instrument Errors," Ph. D. Thesis, Statistics Dept., Southern Methodist University, May 1969. Available as Tech. Report 35, THEMIS, Statistics Dept., Southern Methodist University, Dallas, Texas.

INFORMATION IN TWO-STAGE PROGRAMMING UNDER UNCERTAINTY

David P. Baron

Northwestern University

ABSTRACT

The random variables in two-stage programming under uncertainty are generally treated in a passive manner in that no information regarding the random variables or the process generating the random variables may be obtained. This paper develops the economics of information for the case in which the probability distributions are discrete. A multinomial process is assumed to generate the random variables, and the parameter vector of that process is assumed to be unknown. A Dirichlet prior distribution on the parameter vector is used, and the computation of the value of information thus involves a Dirichlet-multinomial distribution on the random variables. An example involving producing to meet uncertain demands is presented.

I. INTRODUCTION

The use of linear programming for the allocation of resources and the scheduling of activities is limited by the knowledge the decision maker has regarding the values of the coefficients in the program. Early approaches to programs involving uncertainty provided "fat" or "slack" in the program to compensate for the uncertainty [10]. One formal method for explicity incorporating uncertainty is two-stage linear programming under uncertainty originally developed by Dantzig [3] and studied by Dantzig and Madansky [4], El Agizy [6], Madansky [9], and Wets [15]. The form of the program considered in this paper is

$$\text{(1)} \qquad \begin{aligned} &\text{maximize } z = E_d[cx - c_0 w - c_u y] \\ &\text{subject to } \quad Ax \leqslant b \\ &\qquad\qquad\quad Tx + Iy - Iw = d \\ &\qquad\qquad\quad x \geqslant 0,\ y \geqslant 0,\ w \geqslant 0, \end{aligned}$$

where c and x are n-component vectors, y, w, c_0, c_u, and d are J-component vectors, T is a Jxn matrix, I is the identity matrix, b is an m-vector, A is an mxn matrix, and E_d denotes expectation with respect to the random vector d. The program may be thought of as selecting a production activity, x, in light of a random demand, d, and later compensating for any divergence from the actual demand. The sequence proceeds as follows: (1) an activity level x is selected, (2) the random demand is observed, and (3) the decision maker makes up any shortage by purchasing y_i units at a cost of c_{u_i} per unit or disposing of w_i units at a cost of c_{0_i} per unit. Alternatively, at the second stage a shortage cost, c_u, for $y = (d - Tx) > 0$ or a holding cost, c_0, for $w = (Tx - d) > 0$ is incurred. In multicommodity inventory problems, T may equal I. Ferguson and Dantzig [8], [2, ch. 28] have used such a program in the allocation of aircraft to routes where demand is uncertain. The two-stage program considered herein is a special case of a class of programs with recourse [14–16]. In this case, the recourse is linear, but the concepts developed

in this paper apply to more general formulations. The concepts may also be generalized to programs with more than two-stages, but the computations become more complex.

The uncertain demand in two-stage programs is usually treated in a passive manner in that no information regarding the future level of demand or the stochastic process generating demand can be obtained. This paper considers the economics of acquiring such information for the case in which the probability distribution of demand is discrete. Avriel and Williams [1] have recently developed upper bounds on the value of perfect information in programs with recourse. The approach presented here permits the direct determination of the value of sample information. The basic concepts introduced here may be extended to the case of continuous distributions, but the results are not as tractable. The probability model is developed in section II, and the economics of information is considered in section III. An example is presented in section IV, while conclusions are offered in the last section.

II. THE PROBABILITY MODEL

Consider an element, d_i, of the demand vector, and assume that d_i may take on values d_{ik}, $k=1, \ldots, K_i$. Let the probability of d_{ik} occurring be denoted by p_{ik}. A multinomial process may be used to represent the uncertain demand, with the indicator variable, r_{ik}, denoting the occurrence ($r_{ik}=1$) or nonoccurrence ($r_{ik}=0$) of d_{ik}. The likelihood function for a sample of size N_i from such a process is

$$f(r_i \mid p_i)=\frac{N_i!}{r_{i1}!r_{i2}! \ldots r_{iK_i}!} p_{i1}^{r_{i1}} p_{i2}^{r_{i2}} \ldots p_{iK_i}^{r_{iK_i}}, \tag{2}$$

where $r_i=(r_{i1}, \ldots, r_{iK_i})$ and $p_i=(p_{i1}, \ldots, p_{iK_i})$.

In general, the vector parameter, p_i, may be unknown, and the decision maker is assumed to assess a prior distribution $f(p_i)$ on the unknown parameter. The revision of the prior distribution in light of sample information via Bayes' rule is simplified if the prior is a member of a family that is conjugate to the multinomial process [12] [5]. A brief development will be presented to indicate this simplification.

Consider a prior density of the Dirichlet (vector beta) family,†

$$f(p_i)=\frac{\Gamma(m_{i1}+m_{i2}+\ldots+m_{iK_i})}{\Gamma(m_{i1})\Gamma(m_{i2}) \ldots \Gamma(m_{iK_i})} p_{i1}^{m_{i1}-1} p_{i2}^{m_{i2}-1} \ldots p_{iK_i}^{m_{iK_i}-1}, \tag{3}$$

where m_i is the parameter vector of the distribution. Bayes' rule for combining a sample result r_i with the prior $f(p_i|r_i)$ is

$$f(p_i|r_i)=\frac{f(r_i|p_i)\cdot f(p_i)}{\int_{S_i} f(r_i|p_i)\cdot f(p_i)\,dp_i},$$

where the integral in the denominator is over the K_i-1 dimensional unit simplex denoted by S_i. The denominator is a constant, so the posterior is proportional to the numerator,

$$f(p_i|r_i) \propto f(r_i|p_i)\cdot f(p_i),$$

†The development here assumes that the demands d_i, $i=1, \ldots, J$, are independent. Such an assumption is not required for the two-stage program or for the probability model, but independence considerably simplifies the computation of the value of information. A joint prior distribution on the vectors p_i, $i=1, \ldots, J$, of the matrix beta family [11] may be used. The joint likelihood function is then used in Bayes' rule to obtain a posterior distribution. For the program in (1) only marginal distributions are required, and they are of the form developed in section II. The distribution theory becomes more complex without independence and obscures the concepts involved, so the independence assumption will be made.

where $\propto$ denotes proportionality. Note that

$$f(r_i \mid p_i) \propto p_{i1}^{r_{i1}} p_{i2}^{r_{i2}} \cdot \cdot \cdot p_{iK_i}^{r_{iK_i}}$$

and

$$f(p_i) \propto p_{i1}^{m_{i1}-1} p_{i2}^{m_{i2}-1} \cdot \cdot \cdot p_{iK_i}^{m_{iK_i}-1}.$$

The posterior is then proportional to

$$f(p_i \mid r_i) \propto p_{i1}^{r_{i1}+m_{i1}-1} \cdot p_{i2}^{r_{i2}+m_{i2}-1} \cdot \cdot \cdot p_{iK_i}^{r_{iK_i}+m_{iK_i}-1}, \tag{4}$$

which is a member of the Dirichlet family with parameter vector $t_i = r_i + m_i$. The normalizing constant is

$$\frac{\Gamma(t_{i1} + \cdot \cdot \cdot + t_{iK_i})}{\Gamma(t_{i1}) \cdot \cdot \cdot \Gamma(t_{iK_i})}.$$

Bayes' rule simply involves the replacement of the prior parameter m_i with the posterior parameter $t_i = m_i + r_i$. The Dirichlet distribution is rich in the sense that for some m_i it accurately fits nearly any prior opinions regarding the parameter of the multinomial distribution, so there is little loss of generality in using a prior of that family.

The distribution to be used in the program (1) is that for the demand in a planning period. Such a distribution is referred to as a predictive distribution, which in the formulation used here is a distribution on the indicator for the demand in the planning period d_i for $N_i = 1$. Let that indicator be denoted by $\hat{r}_i$. The predictive distribution is

$$f(\hat{r}_i \mid t_i) = B_i \int_{S_i} \prod_{k=1}^{K_i} p_{ik}^{t_{ik}-1} \cdot \prod_{k=1}^{K_i} p_{ik}^{\hat{r}_{ik}} dp_i,$$

where the constant is

$$B_i = \frac{\Gamma(t_{i1} + \cdot \cdot \cdot + t_{iK_i})}{\Gamma(t_{i1}) \cdot \cdot \cdot \Gamma(t_{iK_i})} \cdot \frac{1}{\hat{r}_{i1}! \cdot \cdot \cdot \hat{r}_{iK_i}!}.$$

The integral may be evaluated to yield

$$f(\hat{r}_i \mid t_i) = \frac{1}{\hat{r}_{i1}! \cdot \cdot \cdot \hat{r}_{iK_i}!} \cdot \frac{\Gamma(t_{i1} + \cdot \cdot \cdot + t_{iK_i})}{\Gamma(t_{i1}) \cdot \cdot \cdot \Gamma(t_{iK_i})} \cdot \frac{\Gamma(t_{i1} + \hat{r}_{i1}) \cdot \cdot \cdot \Gamma(t_{iK_i} + \hat{r}_{iK_i})}{\Gamma(t_{i1} + \hat{r}_{i1} + \cdot \cdot \cdot t_{iK_i} + \hat{r}_{iK_i})}, \tag{5}$$

which will be called a Dirichlet-multinomial distribution. The probability of $\hat{r}_i = (0, \cdot \cdot \cdot, 0, \hat{r}_{ij} = 1, 0, \cdot \cdot \cdot, 0)$ is

$$f(\hat{r}_{ij}) = \frac{t_{ij}}{\sum_{k=1}^{K_i} t_{ik}}, \tag{6}$$

which is the ratio of the parameter t_{ij} to the sum of the parameters. If $m_{ij} = 0$ for all j, the probability is the number of past successes in $\sum_{k=1}^{K_i} r_{ik}$ trials.

III. THE ECONOMICS OF INFORMATION

The program in (1) is a convex program, so the first objective is to develop a linear program equivalent to (1) that simplifies computations. Following Wets [15] let $v_i = T_i x$, where T_i is the ith row of T. The

optimal value of the second stage after the v_i have been determined in stage one is deterministic and conditional on v_i

$$\sum_{i=1}^{J} (c_{o_i} w_i + c_{u_i} y_i) = \sum_{i=1}^{J} \left\{ \begin{array}{l} c_{o_i}(v_i - d_i) \text{ for } v_i - d_i \geq 0 \\ c_{u_i}\ (d_i - v_i) \text{ for } d_i - v_i > 0 \end{array} \right\}. \tag{7}$$

The expected value of the optimal value of the second stage is

$$E_d\left[\sum_{i=1}^{J} (c_{o_i} w_i + c_{u_i} y_i) \right] = \sum_{i=1}^{J} \left[c_{o_i} \sum_{k=1}^{q_i} (v_i - d_{ik}) f(\hat{r}_{ik}) + c_{u_i} \sum_{k=q_i+1}^{K_i} (d_{ik} - v_i) f(\hat{r}_{ik}) \right], \tag{8}$$

where q_i is such that $d_{iq_i} \leq v_i \leq d_{i,\,q_i+1}$. The $f(\hat{r}_{ik})$ are predictive probabilities from (6) developed from the prior distribution. Adding and subtracting $\left[c_{u_i} \sum_{k=1}^{q_i} (d_{ik} - v_i) f(\hat{r}_{ik}) \right]$ inside the brackets in (8) gives

$$E_d\left[\sum_{i=1}^{J} (c_{o_i} w_i + c_{u_i} y_i) \right] = \sum_{i=1}^{J} \left[\bar{c}_i \sum_{k=1}^{q_i} (v_i - d_{ik}) f(\hat{r}_{ik}) + c_{u_i} (\bar{d}_i - v_i) \right], \tag{9}$$

where $\bar{c}_i = c_{o_i} + c_{u_i}$, $\bar{d} = E(d_i)$, and q_i depends on the magnitude of v_i. Introduce the variables v_{ih}, $i = 1, \ldots, J$, defined by

$$\begin{aligned} v_i &= \sum_{h=1}^{K_i+1} v_{i1} \\ 0 &\leq v_{ih} \leq d_{i1} \\ 0 &\leq v_{ih} \leq d_{ih} - d_{i,h-1},\ h = 2, \ldots, K_i, \\ 0 &\leq v_{i,K_i+1}. \end{aligned} \tag{10}$$

Substituting these variables into (9) and adding stage one gives a program equivalent to (1):

$$\text{maximize } z = \sum_{j=1}^{n} c_j x_j - \sum_{i=1}^{J} \left[\sum_{h=1}^{K_i+1} \left(-c_{u_i} + \bar{c}_i \sum_{k=0}^{h-1} f(\hat{r}_{ik}) \right) v_{ih} \right] - \sum_{i=1}^{J} c_{u_i} \bar{d}_i, \tag{11}$$

$$\begin{aligned} \text{subject to } & \sum_{j=1}^{n} a_{ij} x_j \leq b_i,\ i = 1, \ldots, m \\ & \sum_{j=1}^{n} t_{ij} x_j - \sum_{h=1}^{K_i+1} v_{ih} = 0,\ i = 1, \ldots, J \\ & x_j \geq 0,\ j = 1, \ldots, n \\ & 0 \leq v_{i1} \leq d_{i1},\ i = 1, \ldots, J \\ & 0 \leq v_{ih} \leq d_{ih} - d_{i,h-1},\ h = 2, \ldots, K_i,\ i = 1, \ldots, J \\ & 0 \leq v_{i,K_i+1},\ i = 1, \ldots, J. \end{aligned}$$

where $f(\hat{r}_{i0}) = 0$ is the probability that demand is less than d_{i1}. The program in (11) is linear and may be efficiently optimized with an upper-bounding modification of the simplex algorithm. Let z^* be the optimal value of the objective function in (11) and x^* the optimal activity level.

To introduce the concept of information economics, consider the special case in which $A = \Phi$ and $c_j \geq 0$ for all j. The program in (11) is thus a multicommodity inventory problem. The optimal activity, x^*, may be determined by taking partial differences of the objective function. If the decision

maker has perfect information regarding the demand level, d_{ik}, $v_i = d_{ik}$ and $y_i = w_i = 0$ for all i. Let $Tx^h = v^h$, where h denotes the h-th demand configuration $(\hat{r}_{1h_1}, \hat{r}_{2h_2}, \ldots, \hat{r}_{Jh_J})$. The optimal value of the objective function, denoted by z^{**}, is then $z^{**} = E_{d^h}(cx^h)$. An upper bound on the expected value of information regarding demand is $U = z^{**} - z^* \geq 0$. No more than U would be paid to know the demand levels. Equivalently, U is the expected value of being able to determine the activity level x after, rather than before, demand has occurred. Because of the difference in the formulation of the probability model, Avriel and Williams [1, Eq. (15)] give an upper bound on U which they refer to as the expected value of perfect information *EVPI*. For the probability model in this paper the *EVPI* is not U but is given in Eq. (12) below. For $A \neq \Phi$, z^{**} is determined by finding the optimal solution to (11) for each possible demand and then taking the expectation of the value of the objective function with respect to the probability of each demand occurring.

In general, perfect information regarding demand is not attainable, but sample information may be obtained to gain information about the true values of the parameters p_i, $i = 1, \ldots, J$. Consider a sample $N = (N_1, \ldots, N_J)$, where N_i denotes the sample size for the i-th demand.† Posterior to the sample N, the indicator, $\hat{r}_i$, is given by a Dirichlet-multinomial distribution $f(\hat{r}_i|t_i)$ with parameter $t_i = m_i + r_i$, $i = 1, \ldots, J$, where r_i is the sample result. Denote the optimal value of (11) using the posterior predictive probabilities $f(\hat{r}_{ik}|t_i)$ by $\hat{z}$ and the optimal activity by $\hat{x}$. Let the value $\hat{z}^*$ be the objective function in (11) evaluated at x^* and at the posterior predictive probabilities, $f(\hat{r}_{ik}|t_i)$. The difference $(\hat{z} - \hat{z}^*)$ is referred to as the conditional (on a sample result $r = (r_1, \ldots, r_J)$) value of sample information *CVSI*(r) by Raiffa and Schlaifer [12].

$$CVSI(r) = \hat{z} - \hat{z}^* = \sum_{j=1}^{n} c_j(\hat{x}_j - x_j^*) - \sum_{i=1}^{J} \left[\sum_{h=1}^{K_i+1} \left(-c_{u_i} + \bar{c}_i \sum_{k=0}^{h-1} f(\hat{r}_{ik}|r_i) \right) \cdot (\hat{v}_{ih} - v_{ih}^*) \right],$$

where $\sum_{h=1}^{K_i+1} v_{ih}^* = T_i x^*$. The *CVSI*$(r)$ is conditional on a particular sample result r, and the expected value of sample information (*EVSI*) is the expectation of the *CVSI*(r) with respect to $f(r) = f(r_1) \ldots f(r_J)$ which is given by the product of Dirichlet-multinomial distributions developed from the prior distributions, where $f(r_i)$ is given in (6) for $t_i = m_i$.

$$EVSI(N) = E_r[CVSI(r)].$$

The *EVSI*(N) is the maximum amount that would be paid for a sample of size N. The optimal sample size N is determined by maximizing $[EVSI(N) - C(N)]$ with respect to N, where $C(N)$ is the cost of sampling.

The evaluation of *EVSI* may not be as formidable as one might expect. In a very rough sense computational simplifications may be obtained by ordering [13] the sample results according to $(N_i, 0, \ldots, 0)$, $(N_i - 1, 0, \ldots, 0)$, $\ldots$, $(0, \ldots, 1, N_i - 1)$, $(0, \ldots, 0, N_i)$. The ordering must also be done with respect to the J processes. Using the revised simplex method for computations, the ordering may (hopefully) lead to a number of consecutive sample outcomes that yield the same optimal activity level $\hat{x}$ in determining *CVSI*(r).

The value U gives an upper bound to *EVSI*, but a tighter bound is given by the expected value of perfect information *EVPI*. The *EVPI* is the maximum amount that would be paid for a sample that

†N_i need not be integer valued, since it may be scaled relative $\hat{r}_i$. The sample will be considered an integer in the body of the paper.

revealed the true value of the parameters p_i, $i=1, \ldots, J$. If the parameters $p=(p_1, \ldots, p_J)$ are known, multinomial distributions as in (2) with $N_i=1$ are used to develop the predictive probabilities for (11). Let $x(p)$ be the optimal activity as a function of p and $z(p)$ the value the objective function in (11) evaluated at $x(p)$ and $f(\hat{r}_{ik}|p_i)$, $i=1, \ldots, J$. The *EVPI* is then given by‡

$$(12) \qquad EVPI=\int_{S^*} z(p)f(p)dp-z^*=\int_{S_1}\cdots\int_{S_J} z(p)f(p_1)\cdots f(p_J)dp_1\cdots dp_J-z^*,$$

where $S^*=S_1X\ldots XS_J$. The evaluation of *EVPI* is difficult because of the difficulty in constructing the function $z(p)$, but a Monte Carlo sampling technique may be used to generate information that may be used to approximate $z(p)$. That is, randomly choose values of p and solve (11) to obtain $z(p)$. Then fit a response surface [2] to the points obtained, and numerically integrate to obtain the *EVPI*.

IV. AN EXAMPLE

This example is based upon one used by Elmaghraby [7]. Consider a firm producing two products each of which is subject to an independent, uncertain demand which may take on discrete values d_{jk}.

Product 1: Demand d_{1k}	Product 2: Demand d_{2k}
50	20
60	25
70	30
	35

The firm must determine the optimal quantity of each product to produce. For every unit by which demand exceeds production a penalty, c_{uj}, per unit is incurred, and a penalty, c^*_{oj}, per unit is incurred when production exceeds demand. For such a problem $T=I$ and $v=x$. The following table presents the unit penalty costs, selling prices of the products, and production costs.

Product	Selling Price, y_j	Production Cost, g_j	c_{uj}	c^*_{oj}
1.....	\$1.35	\$.89	\$1.40	\$.25
2.....	1.50	.75	1.55	.18

The objective function in the form of (11) with $c_{oj}=c^*_{oj}+y_j$ and $c_j=y_j-g_j$ is

$$\begin{aligned} z=&0.46x_1+0.75x_2+1.40v_{11}-(-1.40+3.0f(\hat{r}_{11}))v_{12}\\ &-(-1.40+3.0(f(\hat{r}_{11})+f(\hat{r}_{12})))v_{13}-(-1.40+3.0(1))v_{14}\\ &+1.55v_{21}-(-1.55+3.23f(\hat{r}_{21}))v_{22}\\ &-(-1.55+3.23(f(\hat{r}_{21})+f(\hat{r}_{22})))v_{23}-(-1.55+3.23(f(\hat{r}_{21})+f(\hat{r}_{22})\\ &+f(\hat{r}_{23})))v_{24}-(-1.55+3.23(1))v_{25}. \end{aligned}$$

‡The function $z(p)$ is assumed to be integrable.

The program is subject to the constraints

$$\begin{aligned} 3x_1 + 5x_2 &\leq 480 \\ 4x_1 + 3x_2 &\leq 480 \\ x_1 - 2x_2 &\leq 60 \\ -x_1 + 2x_2 &\leq 60 \\ x_1, x_2 &\geq 0. \end{aligned}$$

The demand d_{ik} is assumed to be generated by a multinomial process $f(r_i|p_i)$ with p_i. $i=1, 2$ unknown parameters. Let the prior distribution $f(p_i)$ be Dirichlet with parameters $m_{11}=2$, $m_{12}=4$, $m_{13}=4$ and $m_{21}=2$, $m_{22}=7$, $m_{23}=7$, $m_{24}=4$.

The distribution on $\hat{r}_{ik}$ is Dirichlet-multinomial:

$$f(\hat{r}_{11}) = 0.2; f(\hat{r}_{12}) = 0.4; f(\hat{r}_{13}) = 0.4$$
$$f(\hat{r}_{22}) = 1; f(\hat{r}_{22}) = 0.35; f(\hat{r}_{23}) = 0.35; f(\hat{r}_{24}) = 0.2.$$

The probabilities $f(\hat{r}_{ik})$ are those used by Elmaghraby, although he did not use the above probability model.

The constraints used by Elmaghraby are not binding for any demand level, so the case in which $A=\Phi$ is relevant. The optimal solution may be determined by taking partial differences. The optimal solution is $x_1^*=70$, $x_2^*=30$ with $z^*=\$35.73$. Since $c_j \geq 0$, $j=1, 2$, the optimal output with perfect information regarding the demands is $\hat{x}_i = d_{ik}$, and the value of the objective function is $z^{**}=\$49.71$. The upper bound on the expected value of information is $U=z^{**}-z^*=\$13.98$, which is the maximum amount that would be paid for perfect information regarding demand. The upper bound U is the expected loss resulting from making the output decisions prior to observing demand or the expected value of being able to defer the decision until demand has been observed. The bound given by Avriel and Williams [1, Eq. (15)] equals \$18.97.

Sample information may reduce, but not eliminate, the expected loss U. Consider a sample of size $N_1=1$, $N_2=0$. The following table gives the EVSI for three values of the right-hand side of the first constraint.

Right-hand side of first constraint	Optimal solution with prior information (x_1, x_2)	Optimal solution with $r_{11}=1$ (x_1, x_2)	Optimal solution with $r_{12}=1$ (x_1, x_2)	Optimal solution with $r_{13}=1$ (x_1, x_2)	EVSI
480	(70, 30)	(60, 30)	(60, 30)	(70, 30)	\$.29
345	(65, 30)	(60, 30)	(60, 30)	(65, 30)	.15
300	(60, 30)	(60, 30)	(60, 30)	(60, 30)	.00

Binding constraints may thus significantly reduce or eliminate the value of information.

V. CONCLUSIONS

This paper has developed values of sample information for a class of two-stage programs under uncertainty with discrete probabilities represented by a multinomial process. In general, the parameters of such a process will be unknown, and a Dirichlet prior distribution may be used to represent the decision maker's prior judgments. From such a model the expected value of sample information is readily determined. The probability model may also be used to revise the probabilities in a sequence

of two-stage models in which the occurrence of a particular demand level in a period is used to revise the distribution of demand for future periods.

REFERENCES

[1] Avriel, M., and A. C. Williams, "The Value of Information and Stochastic Programming," Operations Research *18*, 947–954 (Sept.–Oct. 1970).

[2] Box, G. E. P., "The Exploration and Exploitation of Response Surfaces," Biometrics, Vol. 10 (1954).

[3] Dantzig, G. B., *Linear Programming and Extensions* (Princeton University Press, Princeton, N.J., 1963).

[4] Dantzig, G. B., and A. Madansky, "On the Solution of Two-Stage Linear Programs Under Uncertainty," vol. 1 of *Proc. of Fourth Berkeley Symp. on Math. Stat. and Prob.*, held at the University of California, Berkeley, 1961.

[5] deGroot, M. H., *Optimal Statistical Decisions* (McGraw-Hill, New York, 1970).

[6] El Agizy, M., "Two-Stage Programming Under Uncertainty with Discrete Distribution Function," Operations Research *15*, 55–70 (Jan.–Feb. 1967).

[7] Elmaghraby, S. E., "An Approach to Linear Programming Under Uncertainty," Operations Research, *7*, 208–216 (1959).

[8] Ferguson, A. R., and G. B. Dantzig, "The Allocation of Aircraft to Routes – An Example of Linear Programming Under Uncertain Demand," Management Science *3*, 45–73 (Oct. 1956).

[9] Madansky, A., "Dual Variables in Two-Stage Programming Under Uncertainty," J. Math. Anal. Appl. *6*, 98–108 (1963).

[10] Madansky, A., "Linear Programming Under Uncertainty," R. L. Graves and P. Wolfe, eds., *Recent Advances in Mathematical Programming* (McGraw-Hill, New York, 1963), pp. 103–110.

[11] Martin, J. J., *Bayesian Decision Problems and Markov Chains* (John Wiley & Sons, Inc., New York, 1967).

[12] Raiffa, H., and R. Schlaifer, *Applied Statistical Decision Theory* (Harvard Business School, Boston, 1961).

[13] Swanson, L. W., and L. G. Woodruff, "A Sequential Approach to the Feed-Mix Problem," Operations Research *12*, 89–109 (Jan.–Feb. 1964).

[14] Walkup, D. W., and R. Wets, "Stochastic Programs with Recourse," SIAM J. Applied Math. *15*, 1299–1314 (1967).

[15] Wets, R., "Programming Under Uncertainty: The Complete Problem," Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, *4*, 316–339 (1966).

[16] Ziemba, W. T., "Computational Algorithms for Convex Stochastic Programs with Recourse," Operations Research, *18*, 414–431 (May–June 1970).

OPTIMAL CONTROL OF MULTI-CHANNEL SERVICE SYSTEMS*

M. J. Magazine

Department of Industrial Engineering
North Carolina State University
Raleigh, North Carolina

ABSTRACT

D. P. Heyman, M. Sobel, and M. J. Magazine among others have shown existence of an optimal policy for control of single server queuing systems. For queues under periodic review existence of an analogous rule is established for multi-server systems. Formulation as a dynamic programming problem is given and proofs for existence are presented for finite horizon, infinite horizon and average cost criteria.

INTRODUCTION.

Define a policy, $\tilde{\pi}$, for determining the number of servers to have open in a queuing system at some observation point as

$$\tilde{\pi}: (i, k; j_0, j_1, \ldots, j_s),$$

where the i, k represents the presence state of the system, i.e., i customers are present and k servers opened. The sequence, $\{j_l\}$, $l=0, 1, \ldots s$, represents some action to be taken; e.g., it specifies how many servers to open or close, if $j_m \leq i < j_n$ for two indices m and n. A policy is said to be optimal if the specified cost criteria is a minimum.

We consider the set of policies, π, which has the added characteristic that $\{j_l\}$ is a *nondecreasing sequence*. It will be shown that the set of optimal policies for either finite or infinite horizon and averaging cost criteria is contained in the set of policies, π.

PREVIOUS RESULTS AND MODEL CHARACTERISTICS.

It has been established, for single server queuing systems when observations are made at the completion of arrival and departure epochs, that the form of the optimal policy for the operation of the system has intuitive appeal. By an optimal policy we mean the specification as to when the server should be opened or closed as a function of the number of customers in the system so as to minimize some cost function. Heyman [2] has shown that a (j^*, i^*) policy is optimal for the M/G/1 system. This says: open a closed server only when there are at least i^* in the system, and close an open server only when there are no more than j^*. Sobel [5] has shown a similar result for a GI/G/1 system. Magazine [4] has shown that a policy of this form is also optimal for the M/M/1 system under periodic review.

Consider the following modifications of these systems:

(i) Assume there are s servers.

(ii) Assume that the decision points are at equally spaced time intervals.

*Research for this report was sponsored in part by projects THEMIS, ARO–D Contract Number DAHCO 4–68C–0002 and NSF Research Initiation Grant Number GK–5611.

(iii) Assume a Poisson arrival stream with rate λ, and that each of the s servers have independent and identical service distributions which are exponential with rate μ.

(iv) Assume there is a finite system capacity M.

Assumption (ii) of periodic review causes problems in existence which do not arise when decision points are at arrival and service epochs, (see [4]). This is so because in the latter case there can be at most one change of state. However, to ensure that we can imbed our model into a Markov Chain without increasing the state space assumption (iii) is necessary.

The decision process is as follows. The system is observed at times $t = 1, 2, \ldots$ and the state of the system (X_t, k) is noted. $X_t = i$ implies there are i customers in the system at time t, and k denotes the number of open servers. The decision is made to leave k servers open, up to $s-k$ additional servers, or to close up to k servers. The sequence of states forms a Markov Chain and changes of state occur according to the following probabilistic law:

$$p_{ij}^{k} = Pr\{X_{t+1} = j | X_t = i \text{ and there are } k \text{ open servers}\}.$$

Consider now the following cost structure:

(i) A—Constant cost of opening a closed server.

(ii) B—Constant cost of closing an open server.

(iii) C—Cost of operating an open server for one time period.

(iv) K(i)—holding cost, i.e., cost incurred when i customers are observed in the system.

The problem is to show that a policy of form π yields a minium cost, where π is redefined as

$$\pi\colon (i, k; j_0^*, \ldots j_{k-1}^*; i_{k+1}^*; \ldots i_s^*),$$

where the $\{j_l^*\}$ and $\{i_l^*\}$ are nondecreasing sequences with the following interpretation:

If $i_m^* \leq i < i_{m+1}^*$ and $k \leq m$, open $m-k$ servers. If $j_{m-1}^* < i \leq j_m^*$ and $k \geq m$, close $k-m$ servers. If $i \geq i_s^*$ or $i \leq j_0^*$ have all the servers opened or all closed, respectively. If i does not meet any of these conditions keep k servers open.

For the single server this is nothing more than the (j^*, i^*) policy mentioned earlier.

DYNAMIC PROGRAMMING FORMULATION.

Define $\phi_n(i, \alpha, k)$ as the minimum expected cost with i customers present, k servers open with n periods remaining in the horizon, using discount factor α, $0 < \alpha < 1$.

Then the recursive relationships are:

$$(1)\qquad \phi_n(i, \alpha, k) = \min \begin{bmatrix} K(i) + kB & + \alpha \sum p_{ij}^{0} \ \phi_{n-1}(j, \alpha, 0) \\ K(i) + (k-1)B & + C + \alpha \sum p_{ij}^{1} \ \phi_{n-1}(j, \alpha, 1) \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ K(i) + & + kC + \alpha \sum p_{ij}^{k} \ \phi_{n-1}(j, \alpha, k) \\ K(i) + & A + (k+1)C + \alpha \sum p_{ij}^{k+1} \phi_{n-1}(k, \alpha, k+1) \\ \cdot & \cdot \quad \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \quad \cdot \\ K(i) + (s-k)A + & sC + \alpha \sum p_{ij}^{s} \ \phi_{n-1}(j, \alpha, s) \end{bmatrix}$$

where $\phi_0(i, \alpha, k) = K(i)$.

To prove the existence of an optimal policy of the form π it is necessary to make the following reasonable assumption on the holding costs:

(I) $K(i)$ is a convex function, i.e.,

$\theta K(j)+(1-\theta)K(i) \geq K(\theta j+(1-\theta)i)$ for $\theta \epsilon [0, 1]$ and the argument for $K(\cdot)$ integer

EXISTENCE PROOFS FOR π

FINITE HORIZON.

LEMMA 1: $K(i+X_1)-K(i+X_2)$ for $X_1 > X_2$ is nondecreasing in i.

PROOF: The convex property in assumption (I) can be rewritten as $K(i+h)-K(i) \geq \frac{1}{\theta}[K(i+\theta h) - K(i)]$ with $j=i+h$. This must hold for all $\theta \epsilon [0, 1]$ and θh integer. Hence, in particular it must be that,

$$K(i+2)-K(i) \geq 2[K(i+1)-K(i)]$$

or,

$$K(i+2)-2K(i+1)+K(i) \geq 0. \tag{2}$$

Denote by $\Delta^{(j)}k$ the j^{th} difference of $K(i)$. Then

$$\Delta^{(1)}K = K(i+1)-K(i)$$

$$\Delta^{(2)}K = \Delta(\Delta^{(1)}K) = K(i+2)-2K(i+1)+K(i)$$

This is precisely (2) and hence,

$$\Delta^{(1)}K = K(i+1)-K(i) \text{ is nondecreasing} \tag{3}$$

Now for $X_1 > X_2$

$$\Delta^{(1)}[K(i+X_1)-K(i+X_2)] = K(i+X_1+1)-K(i+X_1)-[K(i+X_2+1)-K(i+X_2)].$$

But this is nonnegative, since

$$K(i+X_1+1)-K(i+X_1) \geq K(i+X_2+1)-K(i+X_2)$$

from (3). Hence,

$$K(i+X_1)-K(i+X_2) \text{ is nondecreasing in } i.$$

LEMMA 2: $\sum_{j=0}^{M}[p_{ij}^m - p_{ij}^n]K(j)$ is nondecreasing for $m < n$.

PROOF: $\sum_{j=0}^{M} p_{ij}^m K(j) = E_X E_{y_1} \ldots E_{y_m} K\left(i+X-\sum_{j=1}^{m} y_j\right)$, where X is a random variable denoting the number of arrivals in a time period and y_i a random variable denoting the number of services by the i^{th} server in the time period.

Defining $\sum_{j=1}^{m} y_j = Y_m$, it must be shown that

$$E_X E_{y_1} \ldots E_{y_n}[K(i+X-Y_m)-K(i+X-Y_n)]$$

is nondecreasing for $m < n$. Since $Y_n \geq Y_m$, $K(i+X-Y_m)-K(i+X-Y_n)$ is nondecreasing by Lemma 1. Hence, the expectation is also nondecreasing.

THEOREM 1: *For* n=1 *the optimal policy is of the form* π.

PROOF: Let there be k open servers. Then,

$$\phi_1(i,\alpha,k)=\min\begin{bmatrix} K(i)+Bk & +\alpha\sum p_{ij}^0\ K(j)\\ K(i)+B(k-1)+ & C+\alpha\sum p_{ij}^1\ K(j)\\ \cdot\quad\cdot & \cdot\\ \cdot\quad\cdot & \cdot\\ \cdot\quad\cdot & \cdot\\ K(i) & +\quad kC+\alpha\sum p_{ij}^k\ K(j)\\ K(i)+A & +(k+1)C+\alpha\sum p_{ij}^{k+1}K(j)\\ \cdot\quad\cdot & \cdot\\ \cdot\quad\cdot & \cdot\\ \cdot\quad\cdot & \cdot\\ K(i)+A(s-k)+ & sC+\alpha\sum p_{ij}^s\ K(j)\end{bmatrix}$$

For any two alternatives m and $n(m<n)$,

$$K(i)+(\text{constant})+\alpha\sum p_{ij}^m K(j)\begin{pmatrix}\leqslant\\=\\\geqslant\end{pmatrix}$$

$$K(i)+(\text{constant})'+\alpha\sum p_{ij}^n K(j)$$

or,

$$\sum[p_{ij}^m-p_{ij}^n]K(j)\begin{pmatrix}\leqslant\\=\\\geqslant\end{pmatrix}\textit{constant}.$$

Equality can occur for at most one value of i since the left hand side is nondecreasing by Lemma 2. Hence, the sequence of numbers which determine the action to be taken is nondecreasing in i.

Since the ϕ_n, $n=1, 2, \ldots$ are *piecewise* convex functions it does not follow that the same analysis can be used to establish the result for $n>1$.

LEMMA 3: $\sum_{j=0}^{M}[p_{ij}^m\phi_1(j,\alpha,m)-p_{ij}^n\phi_1(j,\alpha,n)]$ is nondecreasing in i for $m<n$.

PROOF:

$$\phi_1(j,\alpha,m)=K(j)+C_1+\alpha\sum_{j=0}^{M}p_{ij}^{k_1}K(j)$$

and

$$\phi_1(j,\alpha,n)=K(j)+C_2+\alpha\sum_{j=0}^{M}p_{ij}^{k_2}K(j)$$

where k_1 *and* k_2 are the optimal decisions and $k_1\leqslant k_2$.† Now,

$$\sum_{j=0}^{M}\{p_{ij}^m\phi_1(j,\alpha,m)-p_{ij}^n\phi_1(j,\alpha,n)\}=E\{K(i+x-y_m)-K(i+x-y_n)$$
$$+C_3+\alpha\sum_{j=0}^{M}p_{ij}^m K(j+x-y_{k_1})-\alpha\sum_{j=0}^{M}p_{ij}^n K(j+x-y_{k_2})\}.$$

†It seems apparent that if one has less servers open before making the decision, he would not have more after the decision. This is so, and can be shown for each of the several cases of the relative values of k_1, k_2, m and n.

where the appropriate expectations are taken. The difference of the first two terms is nondecreasing by Lemma 1. The difference of the last two is also nondecreasing, again by Lemma 1. This can be seen by writing them in the expected value notation and noting that $k_1 \leqslant k_2$ and $m < n$.

Next, let us introduce the following induction hypotheses: Assume for $n = 2, 3, \ldots, r$

i) $\phi_n(i, \alpha, k)$ has an optimal rule given by π, and

ii) $\sum_{j=0}^{M} [p_{ij}^{k_1}\phi_n(j, \alpha, k_1) - p_{ij}^{k_2}\phi_n(j, \alpha, k_2)]$, for $k_1 < k_2$ is nondecreasing in i.

The following two part theorem will establish the result for the finite horizon.

THEOREM 2: *$\phi_{r+1}(i, \alpha, k)$ has an optimal policy given by π; and*

$$\sum_{j=0}^{M} [p_{ij}^{k_1}\phi_{r+1}(j, \alpha, k_1) - p_{ij}^{k_2}\phi_{r+1}(j, \alpha, k_2)]$$

for $k_1 < k_2$ is nondecreasing in i.

PROOF: To guarantee that the sequences $\{j^*\}$ and $\{i^*\}$ are nondecreasing we must look at the difference of two alternatives m and $n (m < n)$. This difference is

$$\sum_{j=0}^{M} [p_{ij}^{m}\phi_r(i, \alpha, m) - P_{ij}^{n}\phi_r(i, \alpha, n)] + \text{constant},$$

which is nondecreasing in i by the second part of the induction hypothesis. To establish that this can be used, it is necessary to establish the second part of this theorem which is done in precisely the manner of Lemma 3.

Infinite horizon

Next, it shall be shown that a policy of form π is also optimal for the infinite horizon by a development similar to Bellman [1], 129–132.

THEOREM 3: *A π-policy is optimal for the infinite horizon, i.e.,*

$$\phi(i, \alpha, k) = \min \begin{bmatrix} K(i) + Bk & & & + \alpha \sum p_{ij}^{0}\phi(j, \alpha, 0) \\ \vdots & \vdots & \vdots & \\ K(i) & + kC & & + \alpha \sum p_{ij}^{k}\phi(j, \alpha, k) \\ \vdots & & \vdots & \\ K(i) + & sC + (s-k)A & & + \alpha \sum p^{s}\phi(j, \alpha, s) \end{bmatrix} \tag{4}$$

has an optimal policy of the form π.

PROOF: (The proof will be done for an arbitrary state (i, k)). Define $T_j(\phi_{n-1})$ as the cost associated with the jth alternative, $j = 0, 1, \ldots, s$.

For example,

$$T_0(\phi_{n-1}) = K(i) + Bk + \alpha \sum_{j=0}^{M} p_{ij}^{0}\phi_{n-1}(j, \alpha, 0).$$

Then,

$$\phi_n(i, \alpha, k) = \min_{j=0,1,\ldots s} T_j(\phi_{n-1}).$$

Let $j(n)$ = index which yields the minimum of $T_j(\phi_n)$, $j = 0, 1, \ldots, s$.

This implies that

$$\phi_n(i, \alpha, k) = T_{j(n-1)}(\phi_{n-1}) \leq T_{j(n)}(\phi_{n-1})$$

and

$$\phi_{n+1}(i, \alpha, k) = T_{j(n)}(\phi_n) \leq T_{j(n-1)}(\phi_n).$$

Subtracting,

$$|\phi_n(i, \alpha, k) - \phi_{n+1}(i, \alpha, k) \leq \max [|T_{j(n-1)}(\phi_{n-1}) - T_{j(n-1)}(\phi_n)|, |T_{j(n)}(\phi_n) - T_{j(n)}(\phi_{n-1})|]$$

$$\leq \max_{j=0,1,\ldots s} [|T_j(\phi_{n-1}) - T_j(\phi_n)|] \leq \max [|K(i) + kB + \alpha \sum p_{ij}^0 \phi_{n-1}(j, \alpha, 0)$$

$$- (K(i) + kB + \alpha \sum p_{ij}^0 \phi_n(j, \alpha, 0))|, \ldots |K(i) + Ck + \alpha \sum p_{ij}^k \phi_{n-1}(j, \alpha, k)$$

$$- (K(i) + Ck + \alpha \sum p_{ij}^k \phi_n(j, \alpha, k))|, \ldots |K(i) + sC + (s-k)A + \alpha \sum p_{ij}^s \phi_{n-1}(j, \alpha, s)$$

$$- (K(i) + sC + (s-k)A + \alpha \sum p_{ij}^s \phi_n(j, \alpha, s))|] \leq \max_r \alpha \left| \sum p_{ij}^r (\phi_{n-1}(j, \alpha, r) - \phi_n(j, \alpha, r)) \right|$$

$$\leq \alpha \sum p_{ij}^k |\phi_{n-1}(j, \alpha, k) - \phi_n(j, \alpha, k)|.$$

Hence,

$$\max_i |\phi_n(i, \alpha, k) - \phi_{n+1}(i, \alpha, k)| \leq \max_i \alpha \sum p_{ij}^k |\phi_{n-1}(j, \alpha, k) - \phi_n(j, \alpha, k)|$$

$$\leq \alpha |\phi_{n-1}(j, \alpha, k) - \phi_n(j, \alpha, k)| \leq \alpha \max_j |\phi_{n-1}(j, \alpha, k) - \phi_n(j, \alpha, k)|.$$

Define, $U_n = \max_i [|\phi_n(i, \alpha, k) - \phi_{n-1}(i, \alpha, k)|]$.
Then,

$$U_{n+1} \leq \alpha U_n \leq \alpha^2 U_{n-1} \ldots \leq \alpha^n U_1$$

from which we deduce that

$$\sum_{n=0}^{\infty} U_{n+1} \leq \sum_{n=0}^{\infty} \alpha^n U_1 = \frac{U_1}{1-\alpha},$$

and that the $\lim_{n \to \infty} U_n = 0$. Thus we conclude that

$$\sum_{n=0}^{\infty} [\phi_{n+1}(i, \alpha, k) - \phi_n(i, \alpha, k)]$$

converges uniformly and that

$$\phi_n(i, \alpha, k) \to \phi(i, \alpha, k)$$

which was defined in (4). Hence, certainly if some π is optimal for each n, one will also be as $n \to \infty$.

AVERAGE COST CRITERION.

The following theorem can now be proven.

THEOREM 4: *A π policy is optimal for the average cost criterion, i.e., for*

$$\phi(i, k) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} g(i, k) \tag{5}$$

where $g(i, k)$ is the one step or immediate cost.

PROOF: Let $(i, k; j_0^*, j_1^*, \ldots, j_{k-1}^*; i_{k+1}^*; \ldots, i_s^*)_\alpha = P_\alpha^*$ be the optimal policy for $\phi(i, \alpha, k)$. Let $\{\alpha_t\}$, with $\lim \alpha_t = 1$, be a sequence such that $P_{\alpha_t} = P_\alpha^*$ for all t. Since the number of possible poli-

icies is finite ($\overset{t\to\infty}{M}+1$ states) such a sequence and, therefore, P_α^* must exist. Let P be any other policy. Then certainly,

$$\phi_P(i, \alpha_t, k) \geq \phi(i, \alpha_t, k).$$

Hence,

$$\lim_{t\to\infty}(1-\alpha_t)\phi_P(i, \alpha_t, k) \geq \lim_{t\to\infty}(1-\alpha_t)\phi(i, \alpha_t, k).$$

Using the result from Kaufmann and Cruon [3], page 100,

$$\phi(i, k) = \lim_{t\to\infty}(1-\alpha_t)\phi(i, \alpha_t, k),$$

where $\phi(i, k)$ is defined by (5) the desired result is obtained, and the policy which minimizes $\phi(i, \alpha, k)$ also minimizes $\phi(i, k)$.

REFERENCES

[1] Bellman, R., *Dynamic Programming* (Princeton University Press, 1957).

[2] Heyman, D. P., "Optimal Operating Policies for M/G/1 Queuing Systems," Operations Research **16**, 362–382 (1968).

[3] Kaufmann, A., and R. Cruon, *Dynamic Programming* (Academic Press, 1967).

[4] Magazine, M. J., "Optimal Policies for Queuing Systems With Periodic Review," doctoral dissertation, University of Florida, 1969.

[5] Sobel, M., "Optimal Average–Cost Policy for a Queue With Start-Up and Shut-Down Costs," Operations Research **17**, 145–167 (1969).

THE TRANSPORTATION PARADOX*

Wlodzimierz Szwarc
School of Business Administration
University of Wisconsin, Milwaukee
Milwaukee, Wisconsin

ABSTRACT

A paradox arises when a transportation problem admits to a total cost solution which is lower than the optimum and is attainable by shipping larger quantities of goods over the same routes that were previously designated as optimal. That is, falling *total* costs are present in moving to the greater shipment quantities. Necessary conditions for this to occur are established and an algorithm for solving this expanded transportation problem is supplied.

1. INTRODUCTION

Suppose we are given an optimal solution to a classical transportation problem where the transportation costs are nonnegative (or positive if desired). Consider the following question: Is it possible to find a cheaper solution than the optimal one by shipping more along the optimal routes in such a way that no destination will receive less and no origin will ship less of the product? This means that at least one consumer will get more and at least one producer will send more.

Surprisingly enough, the answer may be affirmative. This is shown in section 2. It therefore appears that there are two methods of improving a transportation program.

METHOD 1: Change the shipment routes while keeping the demands and supplies constant (using any of the primal transportation methods).

METHOD 2: Ship more product along the same routes (more precisely: not using new routes).

This does not mean that every transportation program can be improved by one of these methods. Improvement by Method 1 is, naturally, possible when a solution being considered is not yet optimal. Additional improvements can be secured via Method 2 when the transportation costs associated with the shipping routes of this solution satisfy certain conditions.

REMARK 1: This may be regarded as an example of "increasing returns to scale." Here, however, falling total costs are involved. Thus "optimizations" might be terminated (in companies or in classrooms) by contravening or ignoring the condition that "rational" optimizations will always take advantage of such falling total costs (discarding excesses if necessary).

Although the phenomenon is evidently more general, the main object of this paper will concern itself with problems of the form

$$\text{min: } z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}, \tag{1}$$

*The author was formerly a member of the faculty of Carnegie-Mellon University. This work was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Contract NONR 760(24) NR 047–048 with the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

under conditions

$$\sum_{j=1}^{n} x_{ij} \geq a_i \qquad i=1, \ldots, m, \tag{2}$$

$$\sum_{i=1}^{m} x_{ij} \geq b_j \qquad j=1, \ldots, n, \tag{3}$$

and

$$x_{ij} \geq 0 \qquad \text{for all } i, j. \tag{4}$$

Let us introduce the following relations

$$\sum_{j=1}^{n} x_{ij} = a_i \qquad i=1, \ldots, m \tag{5}$$

and

$$\sum_{i=1}^{m} x_{ij} = b_j \qquad j=1, \ldots, n. \tag{6}$$

We will deal in the paper with four problems:

Problem I	— i.e., problem	(1), (2), (3), (4),
Problem II	— i.e., problem	(1), (4), (5), (6),
Problem III	— i.e., problem	(1), (2), (4), (6), and
Problem IV	— i.e., problem	(1), (3), (4), (5).

We will want to stress that, in general, Problem I is not equivalent to either of the three remaining problems whether $\sum_i a_i = \sum_j b_j$ or not.

The intent of this paper is to provide the reader with a simple transportation type explanation of the paradox while leaving the general linear programming case for separate treatment. Another reason why we call attention to this paradox in this form is that it can lead to some confusion and has apparently already done so. (See section 2). We may here note that Problem I (as well as Problems III and IV) was solved in [4] by a so called transportation scheme method. Alternations of Problems I, III, and IV, where relations (2) and (3) are replaced by double inequalities

$$a_i \leq \sum_{j=1}^{n} x_{ij} \leq \bar{a}_i \tag{2'}$$

$$b_j \leq \sum_{i=1} x_{ij} \leq \bar{b}_j, \tag{3'}$$

have been also solved in [4]. References [4] and [5] present the theory of this method and supply all necessary proofs to show that this method leads to an optimal solution within a finite number of iterations. Section 3 contains conditions that exist when it is possible to improve a solution of Problem II by Method 2 by an increase of a certain a_i and b_j. This procedure, however, doesn't always lead, as shown in this section, to an optimal solution of Problem I. In section 4 we define the Transportation Scheme and present the Transportation Scheme Method for solving Problem I. In section 5 we solve all examples given in section 2. In section 6 we prove that if all $u_i + v_j \leq 0$ (where u_i and v_j are dual variables associated with the optimal basic solution of Problem II) then Problems I and II are equivalent which means that no improvement of the optimal solution of Problem II via Method 2 is possible.

2. Examples

EXAMPLE 1: To show the so called paradox, consider a 3×4 transportation problem with the following cost matrix.*

25	10	2	30	10
5	15	20	10	15
100	65	0	2	20
5	12	13	15	

The numbers a_i and b_j are on the right and below the cost matrix, respectively, and the shipment costs c_{ij} are displayed in the cells for each i-j routing possibility.

The "uniquely optimal" transportation program for this example is:

	2	8		10
5	10			15
		5	15	20
5	12	13	15	

and its optimal total cost is \$241.

But now suppose that a_3 and b_1 are each increased by one unit to $a_3=21$, $b_1=6$. A new program with the same shipping routes is

	3	7		10
6	9			15
		6	15	21
6	12	13	15	

which is *cheaper* than the previous one by 2 units.

The presence of a zero shipping cost (origin and destination the same) sometimes seems to attract undue attention in presentations utilizing the previous example. Hence we also provide another example.

EXAMPLE 2: Consider a 4×5 transportation problem

14	15	6	13	14	7
16	9	22	13	16	18
8	5	11	4	5	6
12	4	18	9	10	15
4	11	12	8	11	

*This example is from p. 30 in [1]. In view of Remark 1, above, it may be of interest to note that it was encountered (in the early 1950's) from data supplied by accountants at Creole Petroleum Co., Inc., who were concerned with accuracy of cost data relative to program consequences for transporting oil from the fields to storage, refinery, tankers, etc., in their Venezuela operations. (I am indebted to A. Charnes and W. W. Cooper for giving me this account from their personal experiences.)

and two transportation programs with the same routes:

		7			7
4	6		8		18
		5		1	6
	5			10	15
4	11	12	8	11	

		10			10 = 7 + 3
4	3		11		18
		2		4	6
	8			7	15
4	11	12	11 = 8 + 3	11	

The program on the left side is optimal for the original problem, but the second, using exactly the same routes, is cheaper by 6 units. The cost decreases from 444 to 438.

As these examples show, Problem II is not equivalent to Problem I although $\sum_i a_i = \sum_j b_j$.

Now suppose $\sum_i a_i \neq \sum_j b_j$. One may ask whether Problem I is equivalent to III if $\sum_j a_i < \sum_j b_j$ or to IV if $\sum_i a_i > \sum_j b_j$.

The following example answers these questions.

EXAMPLE 3: Consider Problem I with the following cost matrix and numbers a_i and b_j:

5	8	7	6	5
6	10	5	5	3
7	15	3	16	15
15	21	8	18	4
5	14	10	6	

Here $\Sigma a_i = 27 < \Sigma b_j = 35$. We solve Problem III by the method from [4] and obtain the following optimal solution:

	10			10
			6	6
5	4	6		15
		4		4
5	14	10	6	

There exists, however, a feasible solution to Problem I,

	12			12
			6	6
5	2	8		15
		4		4
5	14	12	6	

which is cheaper than the previous one by 8 units $= 255 - 247$.

Example 3 shows that Problems I and III (as well as IV) are in general not equivalent.

REMARK 2: In his book ([3], Problem 9.26 on page 327) G. Hadley asks the reader to prove a theorem which we may paraphrase as: Show that Problem I is equivalent: a) to Problem IV if $\Sigma a_i < \Sigma b_j$ and, b) to Problem III if $\Sigma a_i > \Sigma b_j$. This is, of course, impossible to show.* But even if we reverse the inequality signs in a) and b) this theorem is still not true as seen from Example 3.

3. CONDITIONS OF IMPROVING A TRANSPORTATION PROGRAM BY INCREASING ONE a_i AND ONE b_j.

Consider an arbitrary basic solution $X = \{x_{ij}\}$ (which may be optimal) of Problem II. Let B be a set of (i, j) of all basic variables x_{ij}.† We will call X a "B basic" solution. It is well known (see, for example [4]) that there always exists at least one collection of $m+n$ numbers $u_1, \ldots, u_m$; $v_1, \ldots, v_n$ such that

$$c_{ij} + u_i + v_j = 0 \qquad \text{for all } (i, j) \epsilon B.$$

Therefore (1) becomes

$$z = \sum_{i,j} c_{ij}x_{ij} = \sum_{i,j} (c_{ij} + u_1 + v_j)x_{ij} - \sum_{i,j} (u_i + v_j)x_{ij} = -\left(\sum_i a_i u_i + \sum_j b_j v_j\right).$$

Assume that:

1) For some (p, q)

$$u_p + v_q > 0. \tag{7}$$

2) There exists a positive number Δ such that the B basic solution to Problem II, where a_p is replaced by $a_p + \Delta$ and b_q by $b_q + \Delta$, will be a feasible solution to that problem.

The total transportation costs for that solution are

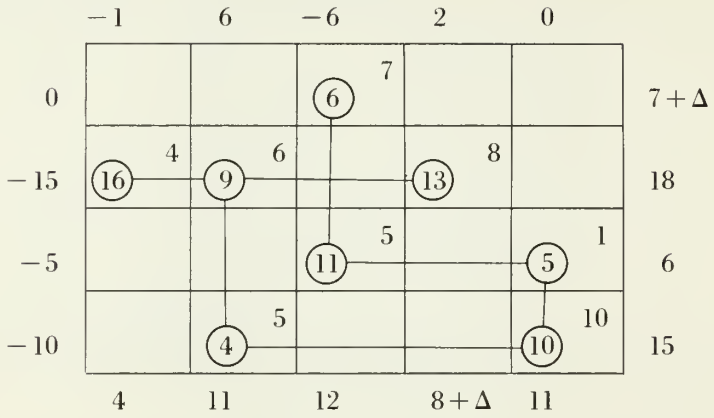
$$z' = z - \Delta(u_p + v_q) < z.$$

We have just shown that (7), as well as the existence of the positive Δ, form a sufficient condition for improving the basic solution by increasing one a_i and one b_j.

* Impossible to show because in Hadley's case (a) we have $\sum_j b_j < \sum_j \left(\sum_i x_{ij}\right) = \sum_i \left(\sum_j x_{ij}\right) = \sum a_i$ which contradicts the assumption $\sum a_i < \sum b_j$. The same contradiction happens for case (b).

† The complement $\bar{B}$ of B may be considered as a collection of all (i, j) for which $x_{ij} = 0$.

To illustrate this, consider the optimal B basic solution $X=\{x_{ij}\}$ of Problem II from Example 2:



The basic elements x_{ij} are written above the corresponding transportation costs c_{ij}, here written inside the circles.*

Numbers u_i and v_j are on the left and above the table.

Here $u_1+v_4=2$. The new B basic solution of Problem II with new a_i and b_j is

		$7+\Delta$		
4	$6-\Delta$		$8+\Delta$	
		$5-\Delta$		$1+\Delta$
	$5+\Delta$			$10-\Delta$

From this result it can be seen that Δ may be selected as any positive number ≤ 5. Taking $\Delta=5$ we obtain the following transportation program:

		12			12
4	1		13		18
				6	6
	10			5	15
4	11	12	13	11	

which is cheaper than the previous one by $5\cdot 2=10$ units.

REMARK 3: Let us return to the optimal B basic solution of Problem II from Example 2. Suppose that there exists a set α of cells $(i, j)\in B$ which have the property that by linking each pair of cells located in the same row and column we obtain a chain consisting of an even number of consecutive perpendicular segments. Let (p, j) and (i, q) be the "endpoints" of the chain.

In Example 2 the chain will appear as shown in Figure 1. Here $\alpha=B-(2, 1)$.

*I.e., we here reverse the conventions for the "stepping stone rules" given in Chap. II of [1]. We do so in order to conform to the conventions in [4].

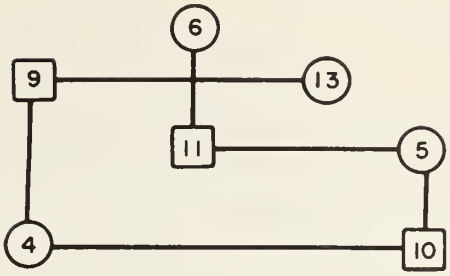


FIGURE 1.

The numbers inside the circles and rectangles are the c_{ij} of the $(i, j)\epsilon\alpha$. Number the cells along the path starting from, say, (p, j) – (in our case $(p, j) \equiv (1.3)$). Divide α into two parts as follows. Assign all odd elements to α_1 and all even elements to α_2. All even elements in Figure 1 are marked by rectangles.

Condition (7) is equivalent to

$$\sum_{(i,j)\epsilon\alpha_2} c_{ij} - \sum_{(i,j)\epsilon\alpha_1} c_{ij} > 0 \tag{8}$$

since as it is easy to show,

$$u_p + v_q = \sum_{(i,j)\epsilon\alpha_2} c_{ij} - \sum_{(i,j)\epsilon\alpha_1} c_{ij}.$$

We have shown in this section that under certain conditions it is possible to improve a solution by increasing a single a_i and b_j.

Suppose that we start with an optimal solution of Problem II and apply this procedure as long as we can, i.e., until a solution is reached which cannot be improved by increasing an a_i and b_j. One may ask whether this final solution is always optimal for Problem I.

The following example shows that the answer is negative. Consider Problem II using the data from Example 3 where instead of $a_2 = 3$ take $a_2 = 11$. Then matrix

$X =$

	5			5
	5		6	11
5	4	6		15
		4		4
5	14	10	6	

is the optimal basic solution to this problem. The corresponding u_i and v_j are written on the left and above the following cost matrix

	0	−8	4	−3
0	5	(8)	7	6
−2	6	(10)	5	(5)
−7	(7)	(15)	(3)	16
−12	15	21	(8)	18

The only positive $u_i + v_j$ are $u_1 + v_3 = 4$ and $u_2 + v_3 = 2$. There are only two different ways of improving X by the mentioned method. The first way (increase a_1 and b_3 by 4) leads from X to

$X_1^* =$

	9		
	5		6
5	0	10	
		4	

while the second way (increase a_2 and b_3 by 4) leads from X to

$X_1^{**} =$

	5		
	9		6
5	0	10	
		4	

and none of these solutions can be further improved by increasing a single a_i and b_j. The optimal solution, however, of Problem I is (see section 5c)

	14		
			6
5		10	
		4	

and it is cheaper than X_1^* (by 10 units) and also cheaper than X_1^{**} (by 18 units)

We summarize the results of this section:

1. Condition (7) as well as the existence of a positive Δ form a sufficient condition to improve a solution via Method 2 by an increase of a certain a_i and b_j.

2. The improvement method mentioned in 1 does not necessarily lead to an optimal solution of Problem I. Let us state at the end of this section the following theorem:

Let X be a B basic solution of Problem II and u_i and v_j be numbers satisfying conditions

$$c_{ij}+u_i+v_j=0 \qquad \text{for all } (i,j)\epsilon B.$$

THEOREM 1: If for all i and j

$$u_i+v_j\leq 0 \tag{9}$$

then Problems I and II are equivalent

The proof of this theorem is based on the transportation scheme method of solving Problem I and has been shown in section 6.

4. THE TRANSPORTATION SCHEME METHOD

By a $m\times n$ transportation scheme (C, M) we mean a problem of the form

$$\text{min: } z=\sum_{i,j} c_{ij}x_{ij}=z_X \tag{1}$$

under constraints

$$\sum_j x_{ij}=a_i \qquad i=1, \ldots, m, \tag{5}$$

$$\sum_i x_{ij}=b_j \qquad j=1, \ldots, n, \tag{6}$$

$$x_{ij}\geqslant 0 \qquad (i,j)\epsilon\phi_1, \tag{10}$$

and

$$x_{ij}\leqslant 0 \qquad (i,j)\epsilon\phi_2, \tag{11}$$

where a_i and b_j are known real numbers with $\sum_i a_i=\sum_j b_j$, $\sum_i |a_i|+\sum_j |b_j|>0$ and where ϕ_1 and ϕ_2 are two disjoint sets of (i, j), such that,

$$\phi_1+\phi_2=\{(i,j)\,|\,i=1, \ldots, m; j=1, \ldots, n\}.$$

Here $C=\{c_{ij}\}$, and $M=(a_1, \ldots, a_m; b_1, \ldots, b_n)$.

Consider two TS problems (C, M) and (C', M) with the same sets ϕ_1 and ϕ_2, where $C'=\{c_{ij}+u_i+v_j\}$ and where u_i and v_j are arbitrary real constants. It is easy to show that problems (C, M) and (C', M) are equivalent.

The "Transportation Scheme" method consists of constructing a sequence of basic solutions of problem (C, M)

$$X_1, X_2, \ldots, X_k \tag{12}$$

with

$$z_{X_1} \geqslant z_{X_2} \geqslant \ldots \geqslant z_{X_k},$$

where X_k is an optimal basic solution.

Paper [4] contains a proof that (12) is a finite sequence.

Let $\bar{X}$ be a B basic solution to problem (C, M) and let $C' = \{c'_{ij}\}$ be defined as follows

$$\text{(13)} \qquad \begin{cases} c'_{ij} = c_{ij} + u_i + v_j \\ c'_{ij} = 0 \qquad \text{for all } (i,j) \epsilon B \end{cases}$$

The optimality criterion, of the transportation scheme (C, M)

$\bar{X}$ is an optimal solution of problem (C, M) if

$$\text{(14)} \qquad \left.\begin{array}{ll} c'_{ij} \geqslant 0 \text{ for all } (i,j)\epsilon\phi_1 & (14') \\ c'_{ij} \leqslant 0 \text{ for all } (i,j)\epsilon\phi_2 & (14'') \end{array}\right\}.$$

REMARK 4: It is easy to see that $\bar{X}$ is an optimal solution to problem (C', M) since $z_{\bar{X}} = 0$ while for any other solution X (basic or not) $z_X \geqslant z_{\bar{X}}$. It is also obvious that $\bar{X}$ is an optimal solution of the problem (C, M) for both problems (C, M) and (C', M) are equivalent.

To solve Problem I we present it as an $(m+1) \times (n+1)$ transportation scheme $(\bar{C}, \bar{M})$ where

$$\text{(15)} \qquad \bar{C} = \begin{pmatrix} c_{11}, \ldots, c_{1n}, & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ c_{m1}, \ldots, c_{mn}, & 0 \\ 0, \ldots, 0, & 0 \end{pmatrix}$$

and $\bar{M} = (a_1, \ldots, a_m, p; b_1, \ldots, b_n, q)$,

where

$$p = \begin{cases} 0 \text{ if } \sum_i a_i = \sum_j b_j \\ \text{a sufficiently large positive number if} \end{cases}$$

$$\sum_i a_i \neq \sum_j b_j$$

$$q = \sum_i a_i - \sum_j b_j + p,$$

and $\phi_1 = \{(i,j) \mid i = 1, \ldots, m;\ j = 1, \ldots, n\} + \{(m+1, n+1)\}$ while ϕ_2 is the set of all remaining cells (in other words: ϕ_2 is the set of all cells in the last row and column except for the cell $(m+1, n+1)$.

The solution procedure of Problem I can be outlined in the following way

STEP 1:

a) If $\sum a_i = \sum b_j$ (Case 1a) solve Problem II and proceed to Step 2.

b) If $\sum b_j - \sum a_i = \Delta > 0$ (Case 1b) find $\min_i \sum_{j=1} c_{ij} = \sum_j c_{rj}$. Solve Problem II where instead of a_r take $a_r + \Delta$. Proceed to Step 3.

c) if $\Sigma a_i - \Sigma b_j = \Delta > 0$ (Case 1c)

$$\min_j \sum_{i=1}^{m} c_{ij} = \sum_i c_{is}.$$

Solve Problem II where instead of b_s take $b_s + \Delta$. Proceed to Step 3.

Let X_1 be a B_1 basic optimal solution of Problem II for either of the cases a, b, c, and u_i and v_j the numbers associated with this solution so that

$$c_{ij} + u_i + v_j = 0 \qquad \text{for all } (i, j) \epsilon B_1.$$

STEP 2: Find $\max_{j:v_j \leq 0} v_j = v_s$ if some $v_j \leq 0$ (Case 2a).

If, however, all $v_j \geq 0$ (Case 2b) find $\max_{i:u_i \leq 0} u_i = u_r$ (at least one of u_i, v_j is ≤ 0).

STEP 3: Determine the set $\bar{B}_1$ of basic variables of the $\bar{B}_1$ basic solution to TS as follows:

$$\bar{B}_1 = B_1 + \{(m+1, s), (m+1, n+1)\} \qquad \text{Case 1c and Case 2a.}$$

$$\bar{B}_1 = B_1 + \{(r, n+1), (m+1, n+1)\} \qquad \text{Case 1b and Case 2b.}$$

The initial $\bar{B}_1$ basic solution $\bar{X}_1 = \{\bar{x}^1_{ij}\}$ of the $(m+1) \times (n+1)$ problem $(\bar{C}, \bar{M})$ has the form

$$\bar{X}_1 = \begin{pmatrix} \bar{X}_1 & \bar{x}^1_{1, n+1} \\ & \vdots \\ \bar{x}^1_{m+1, 1}, \ \ldots & \bar{x}^1_{m+1, n+1} \end{pmatrix}.$$

We are to examine the elements of $\bar{X}_1$ outside the $m \times n$ matrix X_1:

If $\Sigma a_i = \Sigma b_j$ all these elements are zero.

If $\Sigma a_i < \Sigma b_j$ then $\bar{x}^1_{m+1, n+1} = p$, while $\bar{x}_{r, n+1} = -\Delta$ and all remaining elements are zero.

If $\Sigma a_i > \Sigma b_j$ then $\bar{x}^1_{m+1, n+1} = p$, while $\bar{x}_{m+1, s} = -\Delta$ and all remaining elements are zero.

Proceed to Step 4.

STEP 4: Find $\bar{C}_1 = \{\bar{c}_{ij} + \bar{u}_i + \bar{v}_j\}$ $i = 1, \ldots, m+1$; $j = 1, \ldots, n+1$. If $\bar{C}_1$ satisfies (14) then $\bar{X}_1$ is an optimal solution of TS. Otherwise proceed to Step 5.

STEP 5: Apply the procedure given in [4] until encountering a basic solution $\bar{X}_k$ where the associated matrix $\bar{C}_k$ satisfies (14). This solution is optimal. (This procedure will be illustrated in the next section.)

REMARK 5: By removing the last row and column from the obtained optimal solution of TS, we get the optimal solution of Problem I.

5. ILLUSTRATION OF THE SOLUTION TECHNIQUE FOR SOLVING PROBLEM I: EXAMPLES.

a) Let us solve Problem I using the data from Example 1. Here $\Sigma a_i = \Sigma b_j = 45$; therefore, according to Step 1 we solve Problem II and find the following optimal solution

	2	8	
5	10		
		5	15

To determine $\bar{B}_1$, add $\bar{u}_i$ and $\bar{v}_j$ to the rows and columns of $\bar{C}$:

	0	−10	−2	−4	$v_s=0$
0	25	(10)	(2)	30	0
−5	(5)	(15)	20	10	0
2	100	65	(0)	(2)	0
$-v_s=0$	0	0	0	0	0

Here $\max_{j:v_j \leq 0} v_j = 0 = v_1$. Therefore, $\bar{B}_1 = B + \{(4, 1), (4, 5)\} = \{(1, 2), (1, 3), (2, 1), (2, 2), (3, 3), (3, 4), (4, 1), (4, 5)\}$.

Adding to the last row and column $-v_s$ and v_s, respectively (in our case $v_s = v_1 = 0$), we obtain matrix

$\bar{C}_1 = \{\bar{c}_{ij}^1\} = -2$

		2	2	
25	(0)	(0)	26	0
(0)	(0)	13	1	−5
102	57	(0)	(0)	(2)
(0)	−10	−2	−4	(0)

The $\bar{B}_1$ basic solution $\bar{X}_1 = \{\bar{x}^1_{ij}\}$ is as follows:

	2	8			10
5	10				15
		5	15		20
0				0	0
5	12	13	15	0	

Here $p = q = 0$.

From $\bar{C}_1$ we can see that $\bar{X}_1$ is not optimal since in ϕ_2 there exists a positive $\bar{c}^1_{ij}$. Find

$$\max_{(i,j)\epsilon\phi_2} \bar{c}^1_{ij} = 2 = \bar{c}^1_{35}.$$

So element (2, 5) will enter the new basis $\bar{B}_2$.

As known [4] cell (3, 5) together with some cells $(i, j)\epsilon\bar{B}_1$ form a cycle (there exists always exactly one such cycle). See Figure 2.

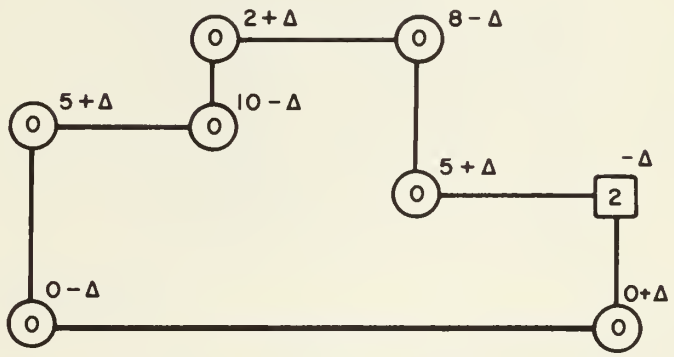


FIGURE 2.

We are looking for a maximal $\Delta > 0$ so that to keep all $\bar{x}^1_{ij} \geq 0$ for $(i, j)\epsilon\phi_1$ and all $x_{ij} \leq 0$ for $(i, j)\epsilon\phi_2$. Here max $\Delta = 8$ and cell (1, 3) is to be removed from $\bar{B}_1$.

Then $\bar{B}_2$ becomes $\bar{B}_2 = \bar{B}_1 - (1, 3) + (3, 5)$. The new $\bar{B}_2$ basic solution $\bar{X}_2$ is

	10			
13	2			
		13	15	−8
−8				8

Adding to $\bar{C}_1$ numbers $\bar{u}_1=\bar{u}_2=\bar{u}_4=0$, $\bar{u}_3=-2$, $\bar{v}_1=\bar{v}_2=\bar{v}_5=0$, $\bar{v}_3=\bar{v}_4=2$, we get the following matrix

$\bar{C}_2=$

25	(0)	2	28	0
(0)	(0)	15	3	−5
100	55	(0)	(0)	(0)
(0)	−10	0	−2	(0)

,

from which it is seen that $\bar{X}_2$ is the optimal solution of the $4\times 5TS$ (since $\bar{C}_2$ satisfies the optimality criterion).

Removing from $\bar{X}_2$ the last row and column, we get the optimal solution of the original Problem I.

	10		
13	2		
		13	15

Observe that the total shipment increased by 8 units.

b) Solve Problem I from the data in Example 2. Here $\Sigma a_i=\Sigma b_j=46$. Proceeding in the same way as before, we obtain the following sequence of basic solutions to the $5\times 6TS$ $\bar{X}_1, \bar{X}_2, \bar{X}_3$ where

$\bar{X}_1=$

		7			
4	6		8		
		5		1	
	5			10	
	0				0

$\bar{X}_2=$

		12			−5
4	6		8		
				6	
	10			5	
	−5				5

and

$\bar{X}_3 =$

		12			−5
4	6		8		
				11	−5
	15				
	−10				10

and where $\bar{X}_3$ is an optimal solution since $\bar{C}_3$ satisfies the optimality conditions. Observe that one may get $\bar{C}_3$ directly from $\bar{C}_1$ by adding $\bar{u}_i$ and $\bar{v}_j$ to obtain zero in the cells of $\bar{B}_3$ (see figure below)

	−2	5	−1	1	0	5
−5	14	15	(6)	13	14	(0)
−14	(16)	(9)	22	(13)	16	0
−5	8	5	11	4	(5)	(0)
−9	12	(4)	18	9	10	0
−5	0	(0)	0	0	0	(0)

The optimal solution of Problem I is

		12		
4	6		8	
				11
	15			

The total shipments increased by 10 units since the sum is equal 56.

c) Let us solve Problem I using the data of Example 3. Here $\Sigma a_i = 27 < \Sigma b_j = 35$. According to Case 1b of Step 1, we find

$$\min_j \sum_{j=1}^{n} c_{ij} = 26 = \sum_{j=1}^{4} c_{2j},$$

and solve Problem II, where we increase a_2 by $8 = 35 - 27$ units. The optimal solution of this problem is as follows:

	5			5
	5		6	$11 = 3 + 8.$
5	4	6		15
		4		4
5	14	10	6	

We set $p = 35\left(= \sum_j b_j\right)$. Then $q = 27$. According to Steps 3 and4, $\bar{B}_1 = B_1 + \{(2, 5), (5, 5)\}$ and $\bar{X}_1$ is as follows:

$\bar{X}_1 =$

	5				5
	5		6	-8	3
5	4	6			15,
		4			4
				35	35
5	14	10	6	27	

where $\bar{x}^1_{25} = -\Delta = -8 = \Sigma a_i - \Sigma b_j$ and $\bar{x}^1_{55} = p = 35$. Proceeding as before, we obtain the following sequence of basic solutions.

$\bar{X}_2 =$

	5			
	9		6	-12
5		10		
		4		
		-4		39

$\bar{X}_3 =$

	14			-9
			6	-3
5		10		
		4		
		-4		39

where $\bar{X}_3$ is optimal. The optimal solution to the original Problem I is given below:

	14		
			6
5		10	
		4	

Here the total amount shipped is equal 39, which is more than $\Sigma b_j = 35 (> \Sigma a_i = 27)$.

REMARK 6: Evidently such rim violations may also be associated with costs, too, as when, for instance, additional overtime authorizations or facilities expansions are involved. To solve Problem I, we deal with a transportation scheme $(\bar{C}, \bar{M})$

$$\bar{C} = \begin{pmatrix} c_{11} & \cdots & c_{1n} & -c_{1,n+1} \\ \vdots & & \vdots & \vdots \\ c_{m1} & \cdots & c_{mn} & -c_{m,n+1} \\ -c_{m+1,1} & \cdots & -c_{m+1,n} & 0 \end{pmatrix},$$

where the nonnegative costs in the last column are taken with the opposite signs since the corresponding x_{ij} are nonpositive.

6. APPENDIX

PROOF OF THEOREM 1: We have to show that

$$\bar{X} = \left(\begin{array}{c|c} X & \begin{matrix}0\\ \vdots\end{matrix} \\ \hline 0 \cdots 0 & \end{array}\right)$$

is an optimal solution of the $(m+1) \times (n+1)$ associated Transportation Scheme Problem. This will imply that X is an optimal solution of Problem I.

Since all $u_i + v_j \leqslant 0$, then either all $v_j \leqslant 0$ or all $u_i \leqslant 0$.

It is sufficient to prove the theorem only for the first case since the proof for the second case is symmetric.

According to Step 2 (of the solution procedure in section 4) we find

$$\max_{j:\, v_j \leqslant 0} v_j = v_s.$$

Using Step 3, determine $\bar{B}_1 = B_1 + \{(m+1, s), (m+1, n+1)\}$ and an initial $\bar{B}_1$ basic solution $\bar{X}$.

Applying Step 4 we find number $\bar{u}_i$ and $\bar{v}_j$ where

$$\left.\begin{aligned} \bar{u}_i &= u_i \qquad i = 1, \ldots, m \\ \bar{u}_{m+1} &= -v_s \\ \bar{v}_j &= v_j \qquad j = 1, \ldots, n \\ \bar{v}_{n+1} &= v_s \end{aligned}\right\}. \tag{16}$$

It is easy to see that matrix $\bar{C}_1 = \{c_{ij} + u_i + v_j\}$ where $\{\bar{c}_{ij}\}$ is defined by (15) satisfies the optimality criterion since a) all $\bar{c}_{ij} \geq 0 (i, j) \epsilon \phi_1$ for X is an optimal solution of Problem II.

b) All $\bar{c}_{ij} \leq 0$ for $(i, j) \epsilon \phi_2$ in view of the fact that

$$\bar{c}_{m+1,j} = v_j - v_s \leq 0 \qquad (v_s = \max_{v_j \leq 0} v_j \geq v_j)$$

$$\bar{c}_{i,n+1} = u_i + v_j \leq 0 \qquad \text{from Assumption (9)).}$$

This completes the proof of Theorem 1.

ACKNOWLEDGMENTS

The author would like to acknowledge his indebtedness to A. Charnes and W. W. Cooper—not only for their help on this manuscript, but also for first calling my attention to the phenomenon of the "Transportation Paradox" during my stay here in the United States in 1962. In fact, I delayed completing this research until I could again return to the United States and only the overly scrupulous reactions of Professors Charnes and Cooper to my invitation prevents them now from appearing as co-authors of this article.

BIBLIOGRAPHY

[1] Charnes, A., and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming*, (J. Wiley and Sons, Inc., New York, 1961), Vol. I, p. 30.

[2] Charnes, A., and W. W. Cooper, "System Evaluation and Repricing Theorems," Management Science *9*, 33–49 (1962).

[3] Hadley, G., *Linear Programming* (Addison Wesley Pub. Comp. Inc., Reading, Mass., 1963), p. 327.

[4] Szwarc, W., "The Transportation Scheme, and Its Applications," Carnegie Institute of Technology, Graduate School of Industrial Administration (Apr. 1963), and Przeglad Statystyczny *11*, 119–167 (1964) (in Polish).

[5] Szwarc, W., "The Transportation Scheme, The Applications of the Transportation Scheme," Bulletin de L'Academie Polonaise des Sciences *10*, 571–576 and 577–581 (1962).

A BACKWARD RECURSIVE TECHNIQUE FOR OPTIMAL SEQUENTIAL SAMPLING PLANS

A. W. Wortham

Texas A&M University

and

E. B. Wilson

Major, USAF

ABSTRACT

This paper presents a procedure akin to dynamic programming for designing optimal acceptance sampling plans for item-by-item inspection. Using a Bayesian procedure, a prior distribution is specified, and a suitable cost model is employed depicting the cost of sampling, accepting or rejecting the lot. An algorithm is supplied which is digital computer oriented.

I. INTRODUCTION

Acceptance sampling by attributes has the acknowledged disadvantage of requiring some manager to answer such a question as follows: What fraction of defective items should be allowed per lot of material delivered from the manufacturing process and what percent of the time should lots exceeding this proportion defective be accepted?

The economic impact of this question has been considered since the very first use of acceptance sampling. Early attempts to minimize sampling cost were concentrated toward reducing expected sample size and thereby reducing inspection cost. Double sampling, sequential sampling, continuous sampling and chain sampling resulted from these attempts. More recently attention has been focused on the costs associated with the accept-reject decisions. Some quality control situations are such that the cost of each defective piece in an accepted lot can be assessed very accurately. On the other hand, cases where defectives could result in injury to people or property make it impossible to assess the cost of accepting a defective. Cost estimates for rejected lots are readily obtained by considering the rectification method following rejection.

Economic models relating the associated costs of acceptance and rejection for single sampling models are reported by numerous authors [4], [5], [8]. A very significant contribution to the theory of optimal sampling plans was presented by Hald [5]. Hald's paper summarizes most of the important concepts concerning the design of optimal single sampling plans. All of this research is characterized by the necessity to rely upon the Bayesian approach to the distribution of defectives in the sample. The Bayesian approach requires that a prior distribution of defectives in the lot be assumed. To avoid proclaiming a prior distribution Breakwell [2], Strange [9], and Van Der Waerden [10] have employed *minimax* procedures to minimize the maximum possible loss. These procedures are particularly objectionable in that the lot fraction defective, p', which yields the maximum loss may be outside the range of possibilities or occur with a very remote frequency.

Champernowne [3] used a linear approximation to determine optimal accept-reject barriers for an item-by-item sequential sampling plan. These barriers converged on a point where further sampling is economically unjustifiable. Lindley and Barnett [6], using the beta prior distribution, derived relationships for finding the sequential sampling accept-reject barriers in terms of the parameters of the beta distribution. Their procedure was geared to a particular sampling cost model.

The method to be reported here is based upon Bellman's principle of optimality and the Markovian property of sequential sampling plans. The technique is computer oriented and provides complete flexibility in the choice of the model and the prior distribution.

II. MODELING CONSIDERATIONS

Consider the following symbols:

N = Lot size,

n = accumulated number of items sampled,

x = accumulated number of defectives observed in the n items sampled,

X = number of defectives in the lot,

$y = X - x$ = number of defectives remaining in the $N - n$ items not sampled,

$E(y|x)$ = expected number of defectives in the unsampled portion of the lot given x defectives in the sampled portion,

c_s = cost of sampling and inspecting one item,

c_v = amount invested in each item,

c_r = cost of replacing, reworking or repairing a defective item, and

c_a = cost attributed to one defective item in an accepted lot.

Wortham and Mogg [12] report nine possible rectification schemes which can be considered in determining average outgoing quality (AOQ). These same possibilities influence the configuration of the cost model. Six of these possibilities are included here to illustrate the model building phase of this technique. Consider the following three sample disposition and two rejected lot dispositions:

Sample;

i) Discard the entire sample,

ii) Replace, repair, or rework all defectives, or

iii) Only enumerate defectives.

Rejected Lots;

i) Discard the entire lot, or

ii) Replace, repair, or rework all defectives.

Let $K_a(n, x)$ be the expected cost of the decision to accept a lot after n items are sampled and x are found defective and let $K_r(n, x)$ be the expected cost of the decision to reject.

Table 1 shows the variety of models arising from the six possibilities. These six possibilities are in no way intended to be exhaustive of the situations that may arise.

III. RECURSIVE RELATIONSHIPS FOR DISTRIBUTIONS

Let $s_n(x|X)$ be the distribution of defectives, x, in a sample of size n given X defectives in the lot, and let $f_N(X)$ be the distribution of defectives in the lot of size N. It follows that $g_n(x)$ the unconditional distribution of defectives in the sample is

$$g_n(x) = \sum_{X=x}^{N-n+x} s_n(x|X) \cdot f_N(X). \tag{1}$$

TABLE 1. *Lot Disposition*

Sample dispositions	(i)	(ii)
(i)	$K_a(n, x) = (c_r + c_s)n + c_aE(y\|x)$ $K_r(n, x) = c_sn + c_rN$	$K_a(n, x) = (c_r + c_s)n + c_aE(y\|x)$ $K_r(n, x) = c_sN + c_rn + c_rE(y\|x)$
(ii)	$K_a(n, x) = c_sn + c_rx + c_aE(y\|x)$ $K_r(n, x) = c_sn + c_rx + c_rN$	$K_a(n, x) = c_sn + c_rx + c_aE(y\|x)$ $K_r(n, x) = c_sN + c_rx + c_rE(y\|x)$
(iii)	$K_a(n, x) = c_sn + c_ax + c_aE(y\|x)$ $K_r(n, x) = c_sn + c_rN$	$K_a(n, x) = c_sn + c_rx + c_aE(y\|x)$ $K_r(n, x) = c_sn + c_rx + c_rE(y\|x)$

Consider the case of sampling without replacement ($s_n(x|X)$ is the hypergeometric distribution);

$$s_n(x|X) = s_n(x|x+y) = \binom{n}{x}\binom{N-n}{y}\Big/\binom{N}{x+y} \tag{2}$$

and

$$g_n(x) = \sum_{X=x}^{N-n}\left[\binom{n}{x}\binom{N-n}{y}\Big/\binom{N}{x+y}\right]\cdot f_N(x+y). \tag{3}$$

Invoking Bayes Theorem we may write the distribution of defectives in the unsampled portion, y, given x as

$$h(y|x = s_n(x|x+y)f_N(x+y)/g_n(x) \tag{4}$$

$$= \binom{n}{x}\binom{N-n}{y} f_N(x+y)\Big/\left[\binom{N}{x+y} g_n(x)\right]. \tag{5}$$

The expectation is

$$E(y|x) = \sum_{y=0}^{N-n} y\cdot h(y|x)$$

$$= \sum_{y=0}^{N-n} y\cdot\binom{n}{x}\binom{N-n}{y}\cdot f_N(x+y)\Big/\left[\binom{N}{x+y} g_n(x)\right] \tag{6}$$

Notice that the quantity $y\binom{n}{x}\binom{N-n}{y}$ can be algebraically manipulated as follows:

$$y\binom{n}{x}\binom{N-n}{y} = \frac{x+1}{n+1}\frac{(n+1)!}{(x+1)!(n-x)!}\frac{(N-n)(N-n-1)!}{(N-n-y)!(y-1)!}$$

$$= (N-n)\frac{x+1}{n+1}\binom{n+1}{x+1}\binom{N-n-1}{y-1}. \tag{7}$$

Now let $w = y - 1$ and

$$E(y|x) = \frac{(N-n)(x+1)}{(n+1)g_n(x)}\sum_{w=0}^{N-n-1} f_N(w+x+1)\binom{n+1}{x+1}\binom{N-n-1}{w}\Big/\binom{N}{w+x+1}. \tag{8}$$

Comparing Equation (8) with Equation (3), it is seen that the quantity under summation is simply $g_{n+1}(x+1)$; therefore,

$$E(y|x) = (N-n)\frac{(x+1)}{(n+1)}\frac{g_{n+1}(x+1)}{g_n(x)}. \tag{9}$$

Thus, the important expectation of the number of defectives in the unsampled portion of the lot given x defectives in the sampled portion is given in terms of the unconditional distribution of defectives in the sample.

Let $E(p_n|x)$ be the expected proportion defective in the unsampled part of the lot given x defectives in the sampled part. It follows that

$$E(p_n|x) = E\left(\frac{X-x}{N-n} \mid x\right) = \frac{E(y|x)}{N-n}$$

(10)

$$= \frac{x+1}{n+1}\frac{g_{n+1}(x+1)}{g_n(x)}.$$

The quantity represented by Equation (10) can also be regarded as the probability of drawing a defective on the $(n+1)^{st}$ item sampled if x defectives have been observed through the nth item sampled.

Now to develop a recursive relationship for $g_n(x)$ under hypergeometric sampling with any general prior distribution, $f_N(X)$, consider the following definition of $g_{n+1}(x)$:

$g_{n+1}(x) = g_n(x)$. [Probability that the $(n+1)^{st}$ item sampled is not defective] $+ g_n(x-1)$. [Probability that the $(n+1)^{st}$ item sampled is defective]

(11) $= g_n(x)\ [1 - E(p_n|x)] + g_n(x-1) \cdot E(p_n|x-1)$

$$= g_n(x)\left[1 - \frac{x+1}{n+1}\frac{g_{n+1}(x+1)}{g_n(x)}\right] + g_n(x-1)\frac{x}{n+1}\frac{g_{n+1}(x)}{g_n(x-1)}.$$

Cancellation and rearrangement of appropriate terms yields

(12) $$g_n(x) = \frac{n+1-x}{n+1}g_{n+1}(x) + \frac{x+1}{n+1}g_{n+1}(x+1).$$

Equation (12) provides a recursive development of the probabilities of x defectives ($x=0, \ldots, n$) from a sample of size n in terms of the probabilities of defectives in a sample of size $n+1$. Figure 1 shows the transition from $n+1$ to n given by Equation (12).

IV. THE IMPORTANCE OF THE VARIANCE OF $f_N(X)$

The product moment $E(x \cdot y)$ can be obtained in a manner similar to the derivation of $E(y|x)$, equations (6) through (9).

From $E(x \cdot y)$ the covariance can be obtained and shown to be positive, negative or zero according to whether the variance of X, σ_X^2, is greater, less or equal to the binomial variance, $Np(1-p)$. *Here*, p is the mean fraction defective or $p = X/N$. The sign of the correlation between the defectives in the sample and the remaining lot, x and y, is a function of the covariance only. This assures that for prior distributions with variance exceeding the binomial variance, the larger the number of sample defectives the larger the expected number of lot defectives. Therefore, when $\sigma_X^2 > Np(1-p)$, $E(y|x)$ increases with x.

As Mood [7] has pointed out, the only sensible prior distributions to consider are those which have a variance exceeding that of the binomial distribution. Otherwise, sampling either reveals nothing about

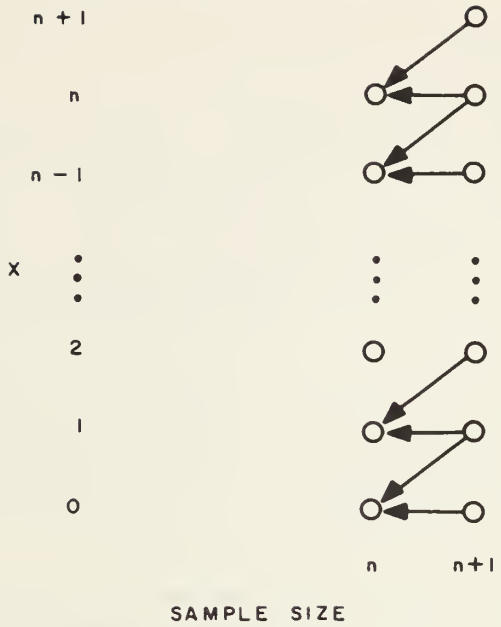


FIGURE 1. Transition for $g_n(x)$.

the lot or a large number of defectives in the sample indicates the opposite about the remaining lot.

When the sample comprises the entire lot, then $g_N(x) = f_N(X)$. Applying Equation (12), $g_n(x)$ can be obtained for any n by the following sequence:

$$g_{N-1}(x) = \frac{N-x}{N} f_N(x) + \frac{x+1}{N} f_N(x+1), \qquad 0 \leq x \leq N-1$$

$$g_{N-2}(x) = \frac{N-1-x}{N-1} g_{N-1}(x) + \frac{x+1}{N-1} g_{N-1}(x+1). \qquad 0 \leq x \leq N-2$$

$$\cdots\cdots\cdots$$

$$g_n(x) = \frac{n+1-x}{n+1} g_{n+1}(x) + \frac{x+1}{n+1} g_{n+1}(x+1), \qquad 0 \leq x \leq n. \tag{13}$$

The system described by relationship (13) can be conveniently executed on a digital computer for any prior distribution $f_N(X)$.

In addition to this recursive development Hald [5] has shown that several prior distributions under sampling without replacement have a quality known as "reproducibility." That is, the unconditional distribution of defectives in samples is identical to the prior distribution. Among prior distributions possessing this trait are the binomial, mixed binomial, hypergeometric, Polya, and uniform. In these cases $g_n(x)$ assumes the identical form as $f_N(X)$, the only difference being n and x replacing N and X.

V. THE BACKWARD RECURSIVE ALGORITHM

Recall that a sequential sampling plan entails the item-by-item inspection of a portion of a lot. Following each item inspected there is a decision to either accept the lot, reject the lot, or inspect another item. The decision is made according to a set of rules derived from Wald's [11] sequential probability ratio. These decision rules are based upon the number of defectives observed through the

current sample. Since Wald's development is based solely upon the risk probabilities, no account is taken of the economic consequences of the decision.

Consider how one might proceed to develop an economically optimal sequential sampling plan. Suppose the decision rules for samples 1 through $n-1$ have been chosen so as to minimize the cost provided by one of the models in Table 1. For the nth sample we wish to select the optimal decision, accept, reject, or sample another item. Let "stage" denote the act of taking a sample, enumerating the total number of accumulated defectives through the latest sample, and making a decision given each of the possible number of accumulated defectives. Let the term "state" designate the number of defectives. Thus at stage n there is an optimal decision for each state zero through n.

By choosing Min $[K_a(n, x), K_r(n, x), K_s(n, x)]$ the optimum decision for each state, x, is accept, reject, or sample another item, respectively. The models $K_a(n, x)$ and $K_r(n, x)$ have been discussed previously. The cost model $K_s(n, x)$ is seen to be the expected cost of deferring the decision until at least one more item is inspected. This expected cost can be obtained as follows:

$K_s(n, x) = K^*(n+1, x)$. [Probability that the $n+1^{st}$ item is not defective] $+ K^*(n+1, x+1)$. [Probability that the $n+1^{st}$ item is defective].

Here, $K^*(n+1, x)$ represents the optimal decision for stage $n+1$ and the probabilities may be replaced by $1-E(P_n|x)$ and $E(P_n|x)$, respectively. The difficulty posed by this procedure lies in determining $K^*(n+1, x)$ when stage $n+1$ has not been optimized. The obvious solution is to turn the procedure around in the dynamic programming sense and optimize the $n+1^{st}$ stage first. Now,

$$K_s(n, x) = K^*(n+1, x) \cdot [1-E(P_n|x)] + K^*(n+1, x+1) \cdot E(P_n|x). \tag{14}$$

Choosing a starting stage, n, which will be the largest possible number of items sampled, it is only necessary to identify the minimum cost provided by either accepting or rejecting the lot for each state, x. To provide a feasible solution, the model must possess the following properties:

i) If $K_a(n, x) < K_r(n, x)$, then
$K_a(n, x-1) < K_r(n, x-1)$, and

ii) If $K_r(n, x) < K_a(n, x)$, then
$K_r(n, x+1) < K_a(n, x+1)$.

Examination of the models in Table 1 reveals that they all possess this property for positive costs and for the case where $E(y|x)$ increases as x increases.

Given the above condition, it is only necessary to seek Min $[K_a(n, x), K_r(n, x)]$ for $x = 0, 1, \ldots, x_o$, where x_o is the smallest state such that $K_a(n, x_o) > K_r(n, x_o)$. Now, let $C_a^n = x_o - 1$ and $C_r^n = x_o$, where C_a^n is the maximum number of defectives to allow for acceptance and C_r^n is the minimum number of defectives for rejection. For all states $x \leq C_a^n$, the decision is to accept, and for all states $x \geq C_r^n$, the decision is to reject.

Having optimized the final stage, n, we may proceed to the $n-1^{st}$ stage. It is now possible to compare $K_a(n-1, x)$, $K_r(n-1, x)$, and $K_s(n-1, x)$ for the minimum. Since $K^*(n, x)$ has been obtained, Equation 14 allows us to look back at the cost-consequence of delaying the decision. At stage $n-1$ it is only necessary to evaluate the models for $x = C_a^n$. Figure 2 shows that if the optimal decisions at stage n are C_a^n and C_r^n where $C_a^n + 1 = C_r^n$, then at stage $n-1$ state $x \leq C_a^n - 1$ can only lead to acceptance if the decision is delayed until stage n. Likewise, if $x \geq C_r^n$ at stage $n-1$, delaying the decision could only lead to rejection at stage n.

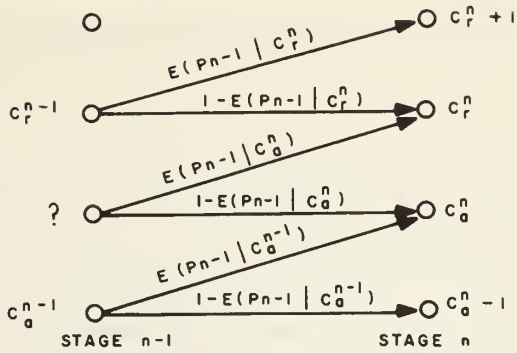


FIGURE 2. State transformation from stage $n-1$ to stage n.

For stage $n-1$, the optimal decision for states $x \leq C_a^n - 1$ will be accept and for states $x \geq C_r^n$ the optimal decision will be reject. For any stage $i \leq n-1$, it is only necessary to examine $K_a(i, x)$, $K_r(i, x)$, and $K_s(i, x)$ for $x = C_a^{i+1}, C_a^{i+1}+1, \ldots, C_r^{i+1}-1$. The optimal decision for $x < C_a^{i+1}$ is accept and for $x \geq C_r^{i+1}$, reject.

It is clear that the procedure outlined above gives the optimal accept-reject barriers in the n, x plane when the following two principles are recalled:

Bellman's Principle of Optimality: "An optimal policy has the property that whatever the initial state and decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision." [1]

Markovian Property: "A process has the Markovian property if the conditional probability of any future event given any past event and the present state of the process is independent of the past event."

Concerning Bellman's principle, the process discussed above starts with an optimal policy for each state of the nth stage and optimal policies are arrived at for all stages subsequently considered based upon these results.

At any stage, n, the only concern is the number of defectives, x, and not in what stage that they may have occurred. The probabilities of zero or one defectives on the $n+1^{st}$ sample is in no way influenced by the individual results of the first through nth sample. Thus the sequential procedure is seen to satisfy the Markovian property.

The complete algorithm can now be summarized as follows:

1. Start the operation at stage "n," where n is either:
   i) The optimal sample size for a single sampling plan,
   ii) An arbitrary upper bound (constraint) on the number of items to inspect, or
   iii) The lot size, $n = N$.
2. For stage n:

i) If $n \neq N$, $g_{n+1}(x)$ and $g_n(x)$ must be obtained from one of the means discussed in section III or IV. If $n = N$, Equation (13) can be employed.

ii) Find,

$$K^*(n, x) = \text{Min}\,[K_a(n, x), K_r(n, x)] \text{ for } x = 0, 1, \ldots x_0 \text{ where}$$

$$K_a(n, x_0) > K_r(n, x_0),$$

$$= K_a(n, x) \text{ for } x < x_0,$$

$$= K_r(n, x) \text{ for } x \geq x_0.$$

iii) Set $C_a^n = x_0 - 1$ and $C_r^n = x_0$.

3. For stage $n-1$:

i) Save $g_n(x)$, $0 \leq x \leq n$, and compute $g_{n-1}(x)$, $0 \leq x \leq n-1$, using Equation (12).

ii) Find.

$$K^*(n-1, x) = \text{Min } [K_a(n-1, x), K_r(n-1, x), K_s(n-1, x)] \text{ for } x = C_a^n \text{ only}$$
$$= K_a(n-1, x) \text{ for } x \leq C_a^n - 1,$$
$$= K_r(n-1, x) \text{ for } x \geq C_r^n.$$

iii) If $K^*(n-1, C_a^n) = K_s(n-1, C_a^n)$, $C_a^{n-1} = C_a^n - 1$ and $C_r^{n-1} = C_r^n$.

If $K^*(n-1, C_a^n) = K_a(n-1, C_a^n)$, $C_a^{n-1} = C_a^n$ and $C_r^{n-1} = C_r^n$.

If $K^*(n-1, C_a^n) = K_r(n-1, C_a^n)$, $C_a^{n-1} = C_a^n - 1$ and $C_r^{n-1} = Cn_a$.

4. For stage $i = n-2, n-3, \ldots, 1$,

i) Compute $g_i(x)$, $0 \leq x \leq i$, from $g_{i+1}(x)$ according to Equation (12).

ii) Find,

$$K^*(i, x) = \text{Min } [K_a(i, x), K_r(i, x), K_s(i, x)] \text{ for } x = C_a^{i+1}, C_a^{i+1}+1, \ldots, C_r^{i+1}-1$$
$$= K_a(i, x) \text{ for } x < C_a^{i+1}$$
$$= K_r(i, x) \text{ for } x \geq C_r^{i+1}.$$

iii) If $K^*(i, x_0) = K_a(i, x_0)$ but $K^*(i, x_0+1) \neq K_a(i, x_0+1)$, $C_a^i = x_0$.

If $K^*(i, x_0) = K_r(i, x_0)$ but $K^*(i, x_0-1) \neq K_r(i, x_0-1)$, $C_r^i = x_0$.

VI. EXAMPLE SAMPLING PLAN:

Consider a situation where sample defectives are replaced by good items and rejected lots are screened (inspected 100 percent) and all defectives are replaced by good items before delivery. This leads to the model:

$$K_a(n, x) = C_s \cdot n + C_r \cdot x + C_a \cdot E(y|x)$$
$$K_r(n, x) = C_s \cdot N + C_r \cdot x + C_r \cdot E(y|x). \tag{15}$$

Assume the following mixed binomial distribution:

$$f_N(X) = \binom{N}{X}[(.9)(.01)^X(.99)^{N-X} + (.1)(.91)^X(.09)^{N-X}]$$
$$N = 100$$
$$E(X) = 10$$
$$\sigma_X^2 = 738.$$

Let the costs be as follows:

$$C_s = \$0.15$$
$$C_r = \$1.00$$
$$C_a = \$2.00$$

Choosing $n=4$ as the starting stage, Table 2 shows the recursive development of the plan.

TABLE 2

State	$g_n(x)$	$E(y\|x)$	$K_a(n,x)$	$K_r(n,x)$	$K_s(n,x)$
Stage 4					
0	0.8645	0.9607	2.52*	15.96	–
1	0.0352	1.6114	4.82*	17.61	–
2	0.0046	77.3184	157.24	94.32*	–
3	0.0271	87.3487	178.30	105.35*	–
4	0.0686	87.3599	179.32	106.36*	–
	$C_a^4=1$	$C_r^4=2$			
Stage 3					
0	0.8733	0.9772	2.40*	15.98	–
1	0.0287	7.7025	17.00	23.77	12.00*
2	0.0226	87.2387	176.73	104.14*	104.23
3	0.0754	88.2690	180.01	106.28*	–
	$C_a^3=0$	$C_r^3=2$			
Stage 2					
0	0.8829	1.0609	2.42*	16.06	2.51
1	0.0342	43.2232	87.70	59.20	52.62*
2	0.0829	89.0842	180.47	106.09*	–
	$C_a^2=0$	$C_r^2=2$			
Stage 1					
0	0.9000	1.8810	3.91	16.88	3.38*
1	0.1000	82.0710	165.28	98.07	96.94*
No decision this stage.					

The plan resulting from this example is depicted graphically in Figure 3.

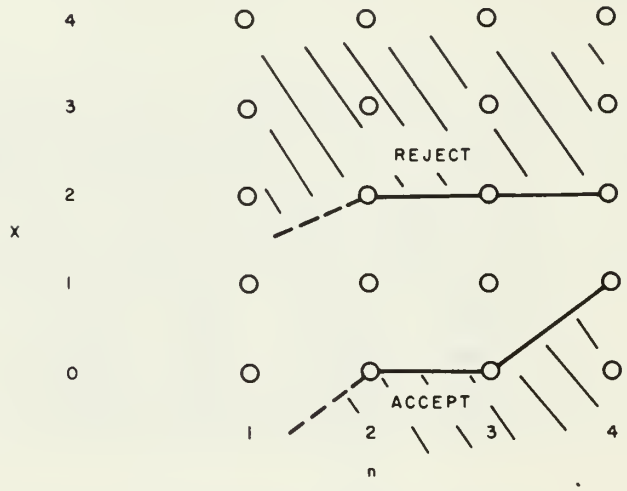


FIGURE 3. Sampling plan.

VII. CONCLUSION

The above algorithm is supplied in a form that can be adapted to a computer code with minimum difficulty. In fact, a code was developed by the authors and found to be quite efficient. Using the WATFOR version of FORTRAN IV on the Texas A&M University IBM 360 computer, the example problem shown above plus another example with twenty-five stages executed in 4.68 seconds. Computing time is seen to be largely a function of the prior distribution. When the functional form of the prior distribution, $f_N(X)$, is included in the program, $f_N(X)$ must be evaluated for each value of X from $X=0$ to $X=N$. Of course this evaluation is only necessary during the Nth stage because Equation (12) can be applied thereafter. With a typical prior distribution such as the mixed binomial evaluation of each point from 0 to N can add significantly to the execution time of the algorithm. If tabulated values of $f_N(X)$ can be supplied and read into the program, execution time can be reduced measurably. Further research on computer code efficiency with this algorithm would be useful in ascertaining its utility as a management tool.

This technique presented above can be employed to design an optimal sequential acceptance sampling plan provided the following conditions exist:

i) The cost of accepting a defective item can be reasonably assessed, and

ii) The nature of the process that assigns defective items to the lot is sufficiently stable and has been studied to the extent that a prior distribution of defectives in the lot can be assumed.

REFERENCES

[1] Bellman, R. E., Dynamic Programming (Princeton University Press, Princeton, New Jersey, 1957).

[2] Breakwell, T. V., "Economically Optimal Acceptance Tests," J. Am. Statis. Assoc. **51**, 243–256 (1956).

[3] Champernowne, D. G., "The Economics of Sequential Sampling Procedures for Defectives," Applied Statistics **2**, 118–130 (1953).

[4] Griffiths, B. and A. G. Rao, "An Application of Least Cost Acceptance Sampling Schemes," Unternehmensforschung **9**, 8–17 (1965).

[5] Hald, A., "The Compound Hypergeometric Distribution and a System of Single Sampling Inspection Plans Based on Prior Distributions and Costs," Technometrics **2**, 275–340 (1960).

[6] Lindley, D. V. and V. D. Barnett, "Sequential Sampling: Two Decision Problems with Linear Losses for Binomial and Normal Random Variables," Biometrika **52**, 507–537 (1965).

[7] Mood, A. M., "On the Dependence of Sampling Inspection Plans Upon Population Distributions," Ann. Math. Statis. **14**, 415–425 (1943).

[8] Pfanzal, J., "Sampling Procedures Based on Prior Distribution and Costs," Technometrics **5**, 47–61 (1963).

[9] Strange, K., "Optimal Sequential Sampling Plans for Known Costs, Minimax Solutions," Unteruchmensforschung **10**, 129–151 (1966).

[10] Van Der Waerden, B. L., "Sampling Inspection as a Minimum Loss Problem," Ann. Math. Statis. **31**, 369–384 (1960).

[11] Wald, A., Sequential Analysis (John Wiley & Sons, Inc., New York, 1947).

[12] Wortham, A. W., and J. M. Mogg, "A Technical Note on Average Outgoing Quality," Quart. J. Quality Technol. **2**, 30–31 (1970).

THE INTERFACE OF COMPUTER SCIENCE AND STATISTICS*

A. F. Goodman

McDonnell Douglas Astronautics Company—Western Division
Huntington Beach, California

ABSTRACT

Current scientific, technical, and management progress is characterized by the generation of a tremendous amount of data for analysis. This, in turn, poses a significant challenge: to effectively and efficiently extract meaningful information from the large volume of data. Two relatively young professions, computer science and statistics, are intimately linked in any response to the challenge. They have consequently become indispensable to scientific, technical, and management progress, occupying a position at its very heart.

Computer science and statistics have each been separately documented by many books as well as numerous papers. However, the interface of computer science and statistics, the area of their interaction, has been documented only in part. This paper begins characterization of the entire interface by providing a structure and an historical background for it.

A structure for the interface is introduced initially, followed by an historical background for the interface presented in two parts. First to be summarized is the evolution of the interface from an interweaving of the mechanical prerequisites to the computer and mathematical prerequisities to computer science and of the foundations for probability and statistics. Development of statistics prior to 1900 then is reviewed.

INTRODUCTION

Current scientific and technical progress is characterized by the generation of a tremendous amount of data for analysis. The philosophy and techniques applied so successfully to natural science and engineering are being adapted to the behavioral and life sciences. A similar trend manifests itself in contemporary progress on the management of large organizations, both private and public. This, in turn, poses a significant challenge: to effectively and efficiently extract meaningful information from the large volume of data.

The objective of statistics is to plan the generation and collection of data and to extract information from data, while the objective of computer science is to utilize the computer for processing data. These two relatively young professions are intimately linked in any response to the challenge. They have consequently become indispensable to scientific, technical, and management progress, occupying a unique position at its very heart.

Computer science and statistics have each been separately documented by many books as well as numerous papers. However, the interface of computer science and statistics, the area of their interaction, has been documented only in part.

This paper begins characterization of the entire interface by providing a sketch of its structure and historical background. That sketch is intended to be preliminary and suggestive, not refined and complete.

---

*The interface of computer science and statistics is the topic of a series of symposia originally sponsored by the Southern California Chapter of the American Statistical Association, the Los Angeles Chapter of the Association for Computing Machinery, and other professional organizations in the Los Angeles area. Three symposia have ensued since the initial meeting in February 1967. This survey article and the following articles by M. Juncosa, and A. F. Goodman were motivated by these symposia.

A structure for the interface is introduced initially, followed by an historical background for the interface presented in two parts. First to be summarized is the evolution of the interface from an interweaving of the mechanical prerequisites to the computer and mathematical prerequisites to computer science and of the foundations for probability and statistics. As evidenced by this evolution, the key-punch and tabulation equipment—which signaled the real beginning of computer science—had only begun to appear prior to 1900. Development of statistics prior to 1900 then is reviewed. A survey of computational mathematics prior to 1900, computer science and statistics after 1900, and the relationship between computational capability and statistics is planned for future publication.

It is hoped that an appreciation of the structure and an historical background for the interface will aid in the initiation and continuation of a meaningful dialogue between computer scientists and statisticians. Only through such a dialogue will the potential of the interface, which has barely begun to be exploited, be fully realized.

In order for the paper to be readily understood by both computer scientists and statisticians, certain technical terms are described in nontechnical language. The objective is to introduce these terms smoothly into the text, providing a sufficiently precise definition without an unduly complex discussion. Technical volumes on computer science and statistics, therefore, should be consulted for a more rigorous treatment.

STRUCTURE FOR THE INTERFACE

INTRODUCTION

It is meaningful to view the computer as another link in the chain of extensions for man's capability:

- His ability to work has been extended by a sequence of aids from simple tools and weapons to complex machines.
- His ability to move has been extended by a sequence of aids from the wheel to high-speed ground and air transportation.
- His ability to communicate has been extended by a sequence of aids from primitive implements to radio and television.
- His ability to observe has been extended by a sequence of aids from the magnifying glass to high-resolution microscopes and telescopes.
- His ability to be intelligent has been extended by a sequence of aids from the abacus to the computer.

Just as the brain provides intelligence required by the former four abilities, the computer may provide intelligence required by an extension of these abilities.

The computer itself is being used for a variety of intelligent activities. In order of increasing difficulty, they are to sort, count, store and retrieve, perform mathematical operations, perform logical operations, and be creative. Implementation of intelligent activities on the computer, however, requires precise specification of a means for determining:

- What computer operations need to be performed and in which sequence.
- When and under what conditions each computer operation is to be initiated.
- How each computer operation is to be performed.
- When and under what conditions each computer operation is to be terminated.

Computer science is concerned with the computer and with the computer's utilization to process data. Statistics is concerned with planning the generation and collection of data and with extraction of information from data.

Statistics extracts information from data by means of analysis, first obtaining a summary description of the data and then using this description to make an inference concerning the data. The former is called descriptive statistics and the latter is called inferential statistics. Planning the generation and collection of data is called design of experiments as a result of its original use in planning agricultural experiments.

Computer science philosophy and techniques have been employed for statistical purposes since the invention of tabulation equipment for sorting and counting 1890 census data. This employment may be characterized as computerized statistics.

That statistical philosophy and techniques could be utilized in the solution of computer science problems and that potentially great dividends should result from this utilization were perceived by the author during the Spring of 1965. Statistical computer science describes such utilization.

The key position of computer science, statistics, and their philosophy and techniques within scientific, technical, and management progress becomes more evident each day. This may be thought of as computer science and statistics in the modern world.

Thus, the three component parts of the interface are computerized statistics, statistical computer science, and computer science and statistics in the modern world. Computerized statistics has been adequately discussed at professional meetings and documented by books and papers. However, the first professional meetings to provide an integrated coverage of all three component parts were *Computer Science and Statistics: Annual Symposia on the Interface*, held in Los Angeles during 1967, 1968, 1969, and 1970.

A discussion of computerized statistics is followed by one of statistical computer science. Then computer science and statistics in the modern world is described. An historical background for this structure is provided by the discussions on evolution of the interface and development of statistics prior to 1900 which are presented in the following two sections.

COMPUTERIZED STATISTICS

Modern statistics was developing but still in its youth when tabulation equipment was invented in the 1880's. The 19th Century use of tabulation equipment for the most elementary statistical analysis of data, that of sorting and counting, constituted the beginning of computerized statistics.

Since then, computerized statistics has grown at an ever-increasing rate. This growth has been accomplished by the utilization of more and better tabulation equipment, desk calculators, and three generations of computers for the more rapid and complex analysis of larger amounts of data. Computerized statistics currently employs the computer as essentially a larger, faster, and more adaptable tabulator/calculator.

The very existence of computerized statistics implies precise communication of the desired statistical analysis to the computer. This precise communication requires a sequence of instructions which becomes longer and more complicated as the statistical analysis becomes more complex.

A sequence of computer instructions for the performance of an elementary statistical analysis is called a statistical computer algorithm. When computer algorithms are combined and coordinated to provide a more complex statistical analysis, a statistical computer program is formed. A system

of statistical computer programs is an integrated collection of computer programs for accomplishing a sequence of complex statistical analyses.

At *Computer Science and Statistics: Second Annual Symposium on the Interface*, Frank Proschan discussed some effects computer science has had on statistics. It is informative to present a summary of his remarks:

- A statistical table which is readily computed may be replaced by a computer algorithm for its generation.
- Novel representations of random variables are being developed for purposes such as statistical computer algorithms.
- A closed-form procedure for solving a statistical problem is no longer required and it may be replaced by an iterative procedure with its implementing computer algorithm.
- Computer simulation may be utilized to suggest and test statistical conjectures.
- The size of n for which an asymptotic statistical approximation is justified has increased.

These remarks serve as an excellent introduction to those below on generation of statistical tables, performance of statistical computations, computer simulation in statistics, and computer-aided design of experiments. Their objective is to illustrate the many ways in which computer science may be applied to statistics.

A statistical table contains the values for a variable of statistical interest which correspond to specified values for one or more other variables. The simplest use of the computer in statistics is to sort and count the data and to generate statistical tables for summary description of the data. A more complex use of the computer is to generate statistical tables of the probability distributions for appropriate functions of the data as an aid in inference concerning the data. The computer has made it feasible to generate more entries of greater accuracy and precision in more statistical tables than ever before. In addition, previously-generated tables for statistical inference have been checked by computer for accuracy and precision.

The increased facility to generate tables for statistical inference not only has increased the availability of such tables, but also has decreased the need for them. This decrease in need occurs either when the table's computer algorithm may be integrated readily into a computerized statistical analysis or when a desired value for the variable of statistical interest may be generated readily upon demand (for example, by utilizing on-line access to a computer).

Use of the computer to perform statistical computations is also an aid in inference concerning the data. The wide variety of existing statistical computer algorithms, programs, and systems testifies to the increasing use of the computer for this purpose.

When a model (mathematical representation) of a system or process is translated into a simulation computer program (computer representation), the system or process and the effects of various factors upon it may be simulated on the computer. Simulation data is used as a substitute for or compliment of actual data.

Computer simulation may be employed in statistics for two purposes: to generate values of a random variable from a given probability distribution and to aid in the development of statistical procedures and theory. Although the former is generally not considered to be a use of computer simulation, it may be viewed as employing the computer to simulate a random process described by the given probability distribution. The latter utilizes an empirical approach to the solution of an analytical problem, although one normally utilizes an analytical approach to the solution of an empirical problem. This is the use to

which Dr. Proschan referred in his remarks. An example of it is the utilization of computer simulation to determine empirically the probability distribution of a random variable generated by a complex process.

A closed-form procedure for solving a statistical problem may either not exist or be potentially improved upon by an iterative procedure. Due to computer development, the employment and evaluation of an iterative procedure have been feasible for some time and continue to become more so with the passage of time. Four questions regarding an iterative procedure, however, should be considered:

- Under what conditions does it converge?
- How rapidly does it converge?
- Under what conditions does it converge to the proper solution?
- To what extent does it improve upon a closed-form (noniterative) procedure?

It is somewhat surprising that iterative statistical procedures have only begun to be used and that development of a meaningful theory for them has been essentially neglected.

The computer has always been far superior to man in generating and examining a large number of alternatives. When a criterion for the evaluation of these alternatives is communicated to the computer, the computer may be used as an aid in making decisions. One such use within statistics which is recent in its origin is for computer-aided design of experiments.

STATISTICAL COMPUTER SCIENCE

During the Spring of 1965, the author attended a lecture by Arthur Samuel on his checker-playing computer program. The essence of having the computer play checkers is to provide it with effective and efficient rules for the intelligent pruning of a decision or game tree. That this very basic computer science problem could be attacked with statistical philosophy and techniques became quickly apparent. Such a thought evolved first into a recognition of the existence and potential of statistical computer science and then into an appreciation of the full expanse and richness of the interface.

Compared to computerized statistics, statistical computer science is in its infancy. A large number of complex, important, and successful examples of the former may be employed to characterize its present form and to project its future evolution. However, a considerable amount of creative imagination needs to be exercised in order to envision the scope and potential of the latter.

The following discussion of computer hardware and software development, computer simulation, information storage and retrieval, artificial intelligence, and planning and management of computer center resources introduces areas of computer science to which statistics either has or should be applied. It is hoped that the reader will then be stimulated to knowledgeable conjecture regarding broader or deeper applications within these areas and additional applications within other areas.

Computer software is a means by which one communicates to computer hardware what to do and when to do it, how to do it and where to do it, or improves his capability of accomplishing either one. The development of computer hardware and software is at the very heart of computer science. Statistical reliability analysis should be either directly applicable or readily adaptable to the development of reliable computer hardware, and statistical concepts should be quite applicable to the development of accurate and efficient computer software.

Many systems and processes are both complex and important with experimentation regarding them being impossible during their design and difficult or expensive during their analysis. For such systems or processes, a closed-form mathematical solution for output in terms of input usually does

not exist and computer simulation may be employed effectively as a substitute for experimentation during system design and as a complement to it during system analysis and improvement. General approaches to this employment, called complete system design and complete system analysis, have been developed by the author [5]. They are depicted by Figures 1 and 2, which indicate how statistical philosophy and techniques may be utilized in system design and analysis.

In complete system design and complete system analysis:

- Mathematical modeling is applicable to quantitative system design (AB) and to quantitative system analysis (AB).
- Statistical estimation, in general, and regression analysis, in particular, are applicable to system model specification (EB) and to system model estimation (EB).
- Statistical testing, in general, and analysis of variance, in particular, are applicable to system model and simulation data comparison (BF), to system and system simulation data comparison (EF), and to system experimental and simulation data comparison (EF).

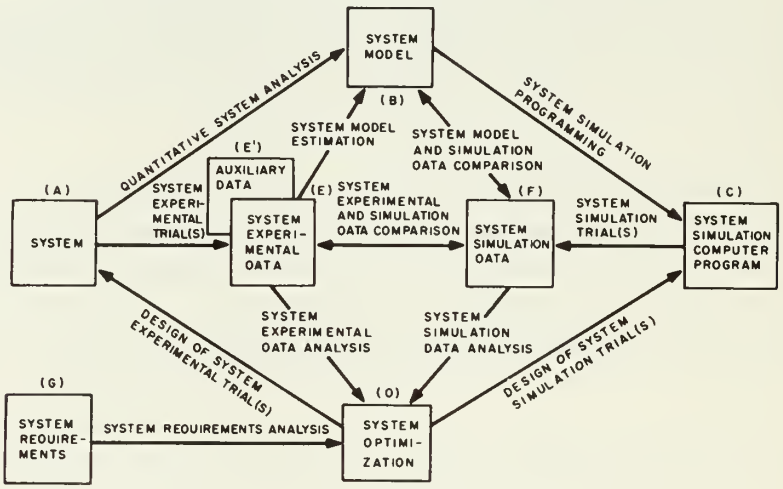


FIGURE 1. Complete system design

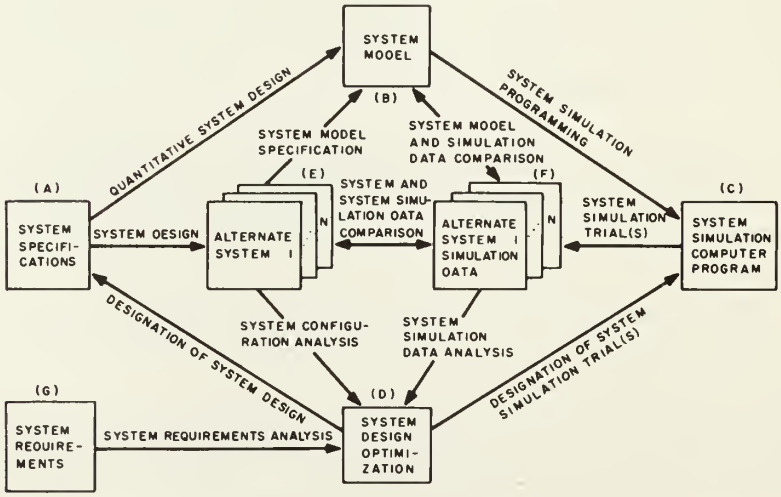


FIGURE 2. Complete system analysis

- Statistical estimation and testing, in general, and regression analysis and analysis of variance, in particular, are applicable to system experimental data analysis (ED) and to system simulation data analysis (FD).
- Statistical design of experiments is applicable to design of experimental trial(s) (DA), to designation of system simulation trial(s) (DC), and to design of system simulation trial(s) (DC).
- Statistical sample survey and analysis are applicable to system requirements analysis (GD).

Computer storage and retrieval—usually, of information describing information—has made it feasible to greatly improve the flow of information in both management and library information systems. Statistics are applicable to computer storage and retrieval, from the initial survey of user requirements through design (see Figure 1) to analysis and improvement (see Figure 2) of the information system. In fact, the author performed a complex statistical analysis of the information requirements for scientists and engineers and proposed goals for the flow of information in Ref. [4].

The objective of artificial intelligence is to employ the computer in creative activity. Since the computer can act creatively only with precise instruction from man, this objective is extremely difficult—some say impossible—to achieve. Progress to date is viewed as either encouraging or discouraging, depending upon one's point of view.

In general, current creative activity on the computer may be divided into pattern recognition, computer-aided instruction, language translation, game playing, problem solving, theorem proving, and musical composition. The basis of this activity is the automatic use of data to make (or, at least, suggest) decisions in the presence of uncertainty, based upon learning from the experience of either man or computer. Intelligent artificial intelligence, therefore, requires statistical philosophy and techniques by definition.

Utilization of mathematics as the primary aid in linguistic analysis may be called mathematical linguistics and utilization of a computer as the primary aid in linguistic analysis may be called computational linguistics. Statistics is highly relevant to both mathematical and computational linguistics.

Computer center resources—machines, men, and money—generally are not planned and managed according to the same standards as the remainder of an organization's resources. Their effective planning and management have, in fact, only begun to be seriously considered. The complexity and importance of this planning and management are increasing with the increasing use of multi-programming, multi-processing, on-line, and time-sharing equipment. In addition, far more information regarding these resources is generated and collected than is meaningfully analyzed.

Statistics may be utilized in the design and selection of computer center resources as it is in complete system design (Figure 1) and in the analysis and improvement of computer center resources as it is in complete system analysis (Figure 2). Furthermore, statistical reliability analysis may be employed throughout design, analysis, and improvement. For financial control, statistics may be applied to the forecasting-budgeting-accounting cycle and for operational control, it may be applied to the predicting-scheduling-monitoring cycle.

COMPUTER SCIENCE AND STATISTICS IN THE MODERN WORLD

Both the thrust of development and the relevance to progress of scientific, technical, and management disciplines evolve dynamically. During any given period, some disciplines are more powerful in their development or more pertinent to general progress than are other disciplines. The thrust and relevance of mathematical science and its branches, always related intimately to scientific progress and now related intimately to technical and management progress as well, also evolve dynamically.

Much of the current action revolves and more of the future action will revolve around computer science and statistics. In addition, this action will influence an ever-increasing sphere of activity in an ever-more significant manner.

A characterization of computer science and statistics in the modern world is presented in terms of their application to management of resources, contract negotiation, and biometrics. These applications may serve to suggest many others, both actual and potential.

Certain games of chance which are favorable to a player and certain aspects of the stock market exhibit similar statistical properties. When the computer is employed to enable the exploitation of these properties in allocating resources, a system for the management of resources in favorable games may be developed.

Incentive contracts for large programs are increasing in number and importance. These contracts are mathematically as well as legally complex. Computer simulation and statistical analysis should be used as an aid in incentive contract negotiation.

Biometrics, the application of quantitative methodology to biological problems, is enjoying a recent revival—sometimes under the title of biomathematics. Implementation of such quantitative methodology is greatly enhanced by the use of computer science and statistics.

EVOLUTION OF THE INTERFACE*

INTRODUCTION

Historically, the interface of computer science and statistics evolved from an interweaving of two strands: development of the mechanical prerequisites to the computer and mathematical prerequisites to computer science and development of probability and statistics. This interweaving constituted only an intersection, but not a real interaction, until the late 19th Century. It is partially the result of the common origin of computer science and statistics in mathematical science and of the past contributions of many mathematicians in multiple areas of mathematical science. The fact that so many high-ranking mathematicians have made significant contributions to both is, however, interesting.

A summary of the evolution which chronologically introduces some of the interweaving between the two strands provides the first portion of an historical background for the interface. The summary depends heavily upon Ref. [16] for its discussion of the major contributors to probability and statistics and upon Refs. [1–3, 6–8, 10, 12–15] for their complementary discussions of mathematicians and mathematical discoveries.

The evolution had an ancient origin. Early intersections occurred in essentially the 16th and 17th Centuries with Pacioli, Cardano, Pascal, Fermat, Huygens, and Leibniz, and later intersections occurred in the 18th and early 19th Centuries with Jacques Bernoulli, DeMoivre, Legendre, Gauss, Laplace, and Boole. Since the late 19th Century, more interactive connections have involved the U.S. Bureau of the Census and Baltimore Department of Health as well as Hollerith and von Neumann.

---

*A preliminary version of this section was published as "The Interface of Computer Science and Statistics: An Historical Perspective" in the June 1968 issue of *The American Statistician*.

ANCIENT ORIGIN

An initial, though tenuous, connection between computer science and statistics may be traced to ancient China, where both the computer and probability have their roots. Around the 6th Century BC, the Chinese invented the bamboo rod abacus. It was almost (if not actually) the first of the mechanical calculators, initial ancestors of the modern electronic computer. In addition, the ancient Chinese may have been the first to realize that the binary system provides the most natural means of representing a number. Sun-Tze, who probably lived during the 1st Century AD (but may have lived as early as the 3rd Century BC), is given credit for being the first mathematician to discuss a problem involving probability. That problem concerned the probability of an expected child being a boy or a girl and might be viewed also as a statistical problem of estimation (Ref. [3], Vol. 1, p. 6; Ref. [10], Vol. 1, p. 516; and Ref. [16], p. 5).

EARLY INTERSECTIONS

The next two intersections, both between the theory of equations and probability, occurred in Renaissance Italy. In 1494, Luca Pacioli published his *Summa de Arithmetica*. This treatise summarized algebra as it then existed and signaled the initiation of a general attack upon the cubic equation by observing that it could not yet be solved algebraically. Pacioli composed a 1494 volume on double-entry bookkeeping as well. He also introduced the first probabilistic problem on gambling into the mathematical literature. The problem involved the equitable division of stakes between two players of unequal skill in a game which is interrupted before its logical conclusion. Girolamo Cardano (Cardan) published his *Ars Magna*, a masterpiece on the theory of equations, in 1545. In addition, he composed a handbook for gamblers, *De Ludo Aleae*, which was published in 1663 as part of his collected works (Ref. [8], p. 68; Ref. [12], p. 64; and Ref. [16], p. 5).

Two 17th Century mathematical astronomers credited with substantial contributions to computational mathematics, Johann Kepler and Galileo Galilei, also made brief reference to probability. However, it is the following connection which might be viewed as the first significant one between computer science and statistics (Ref. [14], p. 129 and Ref. [16], p. 5).

The distinguished 17th Century French mathematician, Blaise Pascal, made considerable achievements in both the development of mechanical calculators and probability. In 1642, he invented the first adding machine with toothed-wheel counters which carried numbers to the next column. This feature made it a major step toward the modern computer. Pascal followed this by his 1654 correspondence with the renowned 17th Century French mathematician, Pierre de Fermat, regarding the probabilistic questions posed to Pascal by the prominent gambler, Chevalier de Mere. Among the questions was the problem first introduced by Pacioli in 1494. In their correspondence, Pascal and Fermat "generalized the problem more and more, until at its close, that which had first appeared as a mere source of perplexity to a gambler had been elevated to a mathematical concept of great import." Fermat also founded the theory of numbers and made great contributions to analytic geometry and the calculus (Ref. [1], p. 57; Ref. [15], pp. 1–3; and Ref. [16], p. 6).

Christiaan Huygens, the well-known Dutch mathematician, followed the Pascal-Fermat correspondence with his small tract, *De Ratiociniis in Ludo Aleae*, in 1657. This tract not only was the first published treatise on the probablistic aspects of games of chance, but also stimulated Jacques Bernoulli and DeMoivre in their subsequent contributions to probability. *Horologium Oscillatorium*, Huygen's book on pendulum clocks in 1673, together with John Wallis' *Arithmetica* represents the

most advanced state of calculus in the period prior to Newton and Leibniz (Ref. [14], p. 142 and Ref. [16], pp. 6 and 7).

Probably the most extensive single connection between computer science and statistics is due to the eminent German mathematician, Gottfried Wilhelm Leibniz. He made substantial contributions to the development of mechanical calculators, computational mathematics, and probability. Leibniz invented a mechanical calculator in 1671 which added, subtracted, multiplied, divided, and extracted roots and was far superior to that of Pascal. A great contribution to computational mathematics was his recognition of advantages of the binary system over the denary system in representing a number. In addition, he and Newton independently crystallized the calculus, an accomplishment of the highest order. The combinatorial analysis of Leibniz contained in his 1666 work called *De Arte Combinatoria* was the link between the primitive beginnings of combinatorial analysis due to Pascal and Fermat and the greatly refined versions of it called the *laws of thought* by Boole in 1854 and symbolic logic by Whitehead and Russell in 1910 (Ref. [1], pp. 118, 119, 123, 125, 434, 446 and Ref. [10], Vol. 1, p. 516).

LATER INTERSECTIONS

A member of the brilliant family of Swiss mathematicians, Jacques (Jacob) Bernoulli, was among the first to develop the calculus significantly beyond that of Newton and Leibniz and to apply it to difficult and important problems. He also made contributions of the highest rank to analytic geometry and the calculus of variations. That so famous and well rounded a mathematician should write *Ars Conjectandi*, the first book devoted entirely to probability and published posthumously in 1713, testifies to the increasing importance and respectability which probability had attained. In addition, it anticipated the concept of inverse probability which is so essential to statistics. It is interesting that Bernoulli and Leibniz, as did Pascal and Fermat before them, exchanged ideas through an extensive correspondence on both combinatorial analysis and the calculus (Ref. [1], p. 133; Ref. [14], p. 165; and Ref. [16], pp. 7, 8, and 11).

With few exceptions, the mathematicians discussed in this paper are known primarily for their contributions to mathematical science, in general, and only secondarily for their contributions to probability or statistics. One such exception is Abraham De Moivre. Although he made contributions to computational mathematics in terms of infinite series and trigonometry, his main contributions were to probability. One such contribution to computational mathematics with which he is generally not credited is Sterling's approximation to factorial x. Although De Moivre's *Doctrine of Chances* of 1718 is essentially a gambler's manual, his *Approximatio ad Summam Terminorum Binomii* $\overline{a+b}^n$ *in Seriem Expansi* of 1733 reveals him as the first conceptual formulator of a law of errors and first discoverer of the normal distribution (Ref. [16], pp. 12–14 and 16).

Had De Moivre's contributions been widely known, probability would have been able to begin its emergence from being solely a foundation for the analysis of games of chance and to grow toward becoming a foundation for the generation, collection, and analysis of data. This, however, had to await the related contributions of Legendre, Gauss, and Laplace. Adrien Marie Legendre first published the method of least squares in 1806, but Carl Friedrich Gauss first formulated it in an unpublished manuscript which was sent to Olbers in 1802. In numerous works on astronomy and science, Gauss also established a firm foundation for the theory of errors from which statistics has descended directly. Pierre-Simon de Laplace through his *Theorie Analytique des Probabilities* in 1812 and *Essai Philosophique sur les Probabilities* in 1814 made one of the greatest contributions to probability. Although

Gauss was Legendre's superior in both the breadth and depth of his mathematics, they both made significant contributions to the theory of numbers, mathematical astronomy, the calculus, and the computation of tables. Laplace made considerable contributions to the calculus and mathematical astronomy—which he called celestial mechanics (Ref. [1], p. 259; Ref. [14], pp. 209 and 211; and Ref. [16], pp. 19, 21, 22, and 24).

Other famous mathematicians who wrote important papers on probability and the theory of errors prior to 1850 were Bessel, Poisson, Fourier, Cauchy, De Morgan, and Tchebecheff. In addition, probability was applied to the computation of mortality tables and annuities by such other well-known mathematicians as Halley, d'Alembert, Euler, and Daniel Bernoulli (Ref. [16], pp. 20 and 30).

The *Laws of Thought* published by George Boole in 1854 embodies in one volume significant contributions to both symbolic logic and probability. This volume devoted approximately two-thirds of its contents to symbolic logic and the remainder to probability, providing a logical and mathematical foundation for both (Ref. [2]).

MORE INTERACTIVE CONNECTIONS

Contemporaneous with and parallel to the construction of a probabilistic basis for statistics was the increase in both need for and collection of statistical data. From these emerged the twin cornerstones of statistics: a mathematical theory of randomness and its treatment and an empirical requirement for the collection and analysis of data (Ref. [16], p. 31).

From its very origin, tabulation equipment has performed a function which is inherently statistical in nature. This origin occurred in the 1880's when an engineer-statistician, Herman Hollerith, was motivated by the U.S. Bureau of the Census to develop a mechanical means for processing 1890 census data. Hollerith was successful in developing the first working models of mechanical tabulation equipment which utilized punched cards. The tabulation equipment is credited with saving the Bureau of the Census $5 million and making the 1890 census results available within a reasonable period of time, which probably would not have occurred without such equipment (Ref. [6], p. 59; and Ref. [8], p. 24).

Prior to 1890, the Baltimore Department of Health used the first electrical tabulation equipment employing punched cards for the handling of vital statistics. Hollerith himself developed an electric sorting machine and an electric key-punch machine by 1900. Hollerith's ideas were developed and perfected by the International Business Machine (IBM) Corporation, while those of his assistant, Powers, were refined by the Remington Rand Company. These industrial protagonists would again take separate paths in the more intense competition which followed the invention of the modern computer (Ref. [3], Vol. 4, pp. 549 and 550; Ref. [6], p. 59; and Ref. [13], pp. 54–57).

To augment this tabulation equipment, there were mechanical and electrical desk calculators which have been refined successively since the close of the 19th Century. The bulk of statistical computations were performed on these desk calculators until quite recently as anyone who was a student in applied statistics during that era can painfully testify.

The most eminent 20th Century mathematician to make substantial contributions to both computer science and statistics was John von Neumann. In 1944, he and Oskar Morgenstern published the *Theory of Games and Economic Behavior.* This book greatly influenced not only statistics, but the social sciences as well. Von Neumann was also one of the pioneers in computer science with such papers as "The General and Logical Theory of Automata" (Ref. [10], Vol. 4, p. 2070; and Ref. [7], pp. 2 and 3).

The first large-scale electromechanical computer was begun in 1939 and became operational in 1944. It was a cooperative effort between IBM and Harvard University and was called the Mark I. Begun

in 1942 and operational in 1946, the University of Pennsylvania's ENIAC was the first large-scale, high-speed electronic computer. The former became the predecessor of IBM's line of electronic computers and the latter became the predecessor of Remington Rand's line of electronic computers. In 1951, Remington Rand delivered the first commercially available, large-scale, high-speed electronic computer—the Univac I—to the U.S. Bureau of the Census. The era of the electronic computer had been launched and a statistical application had participated in that launching (Ref. [3], Vol. 4, p. 550; and Ref. [13], p. 54).

An excellent survey of the development of electronic computers is presented by Ref. [11]. Reference [9] contains a well-written discussion of "Computers as an Instrument for Data Analysis."

DEVELOPMENT OF STATISTICS PRIOR TO 1900

INTRODUCTION

Combined with the section on evolution of the interface, this section provides an historical perspective in which to view the interface meaningfully. The treatment depends almost solely upon Ref. [16]. Contents of this statistical classic probably are not well known to the majority of current statisticians and not known at all to the majority of current computer scientists.

Progress in the collection of data and development of a framework for their analysis occurred independently of each other and essentially in parallel. Adolphe Quetelet fused the need created by the former and the capability provided by the latter to create modern statistics. A discussion of evolution in the collection of data is followed by a description of development of a framework for analysis. Then a treatment of Adolphe Quetelet and one of progress following Quetelet are presented.

COLLECTION OF DATA

The collection of data began as early as recorded history: King David numbered his people and the Egyptians, Babylonians, and Romans kept records of state resources. Pepin the Short in 758 and Charlemagne in 762 required descriptions of church lands and the *Doomsday (Judgement Day) Book* of 1086 listed people, places, and property. A Paris tax record of 1292 contained the names and locations of taxable artisans, while deaths and marriages were registered in early 14th Centruy Burgundy (Ref. [16], pp. 31 and 33).

Sebastian Munster began the investigation into scientific statistics of national resources with his *Cosmographia* of 1536 and 1544. In 1660, the first course of lectures on "Staatskunde" was given by Herman Conring at the University of Helmstedt (Ref. [16], pp. 32 and 33).

John Graunt's *Observations on the London Bills of Mortality* in 1662 marked the start of tabulation and publication of data. For the first time, data were actively operated upon and distributed rather than merely being collected. The first census of modern times to be taken by name was for Canada in 1666 (Ref. [16], pp. 34 and 36).

The purchase price of annuities was scientifically determined for the first time by Jan de Witt in 1671 and the first mortality tables were published by Edmund Halley in 1693. From these beginnings evolved insurance societies as well as both public and private agencies for the tabulation and publication of data (Ref. [16], pp. 36 and 37).

FRAMEWORK FOR ANALYSIS

The existence of organizations dedicated to the tabulation and publication of data, in turn, generated an increasing need for their analysis. As has occurred so often in the history of science, the necessary theoretical framework for the analysis was independently developing for entirely different purposes. Development of the mathematical prerequisites for statistical analysis of data is now reviewed.

In 1773, Abraham De Moivre conceptually formulated a law of errors and discovered the normal distribution. Pierre-Simon de Laplace rediscovered the normal distribution in 1778 and he stated an approximation for the integral of e^{-x^2}, to be used in the computation of tables for it, in 1793. Such tables were published in a 1799 book on refraction by the French physicist, Kramp (Ref. [16], pp. 13, 14, 20, and 58).

The method of least squares was formulated by Carl Frederich Gauss in an unpublished manuscript of 1802 and was published by Adrien Marie Legendre in 1806. In 1808, Robert Adrian investigated the (bivariate) probability distribution of the simultaneous occurrence of two independent errors. Gauss used the method of least squares on planet orbit data in his great work of 1809, *Theoria Motus Corporum Coelestium in Sectionibus Coniciis Solem Ambientium*, and he also inaugurated the study of human errors in observation with his suggestion that reaction time be measured (Ref. [1], p. 259; and Ref. [16], pp. 22, 24, and 94).

Two texts of that era which may be viewed as statistical in nature were Johann Franz Encke's *Uber die Methode der Kleinsten Quadrate* of 1832 and Augustus de Morgan's *An Essay on Probabilities and on Their Application to Life Contingencies and Insurance Offices* of 1838 (Ref. [16], pp. 25 and 26).

Contributions to the investigation of moments for a probability distribution were made by Laplace in 1778, Kramp in 1799, Gauss in 1816, and Encke in 1834. Gauss in 1816 and Encke in 1832 did work related to the percentiles of a probability distribution. Finally, the concept of correlation was almost (but not quite) discovered by Adrain in 1808, Laplace in 1810, (Giovanni Antonio Amedeo) Plana in 1812, Gauss in 1823, and (Auguste) Bravais in 1846 (Ref. [16], pp. 71, 72, 83, 94, and 96).

QUETELET, THE FIRST MODERN STATISTICIAN

The two elements of statistics, collection of data and a framework for analysis, had been developed independently by different people and for different reasons. It remained for a uniquely qualified 19th Century scholar—who perceived the need created by the former and possessed a knowledge of the capability provided by the latter—to fuse them into a single discipline. This man was Adolphe Quetelet: astronomer, historian of science, mathematician, physicist, and the first modern statistician (Ref. [16], p. 39).

Quetelet studied mathematics under (Joseph) Fourier and probability under Laplace. He was director of the Royal Observatory at Brussels, planned the 1829 census of Belgium, and organized the Belgium Commission Centrale de Statistique in 1841. In addition, he instigated the 1834 formation of the London Statistical Society and the First International Statistical Congress of 1853 (Ref. [16], pp. 39 and 40).

The concept of statistics as a general body of techniques applicable to any observation-based science is due to Quetelet and so is the concept of "*l'homme moyen*"—"the average man." In his 1846 *Lettres sur la Theorie des Probabilities Applique aux Sciences Morales et Politiques*, he used "center of gravity" for the arithmetic mean and he also proposed the quantity which Francis Galton

subsequently called the quartile range and currently is called the semi-interquartile range as a measure of dispersion (Ref. [16], pp. 40, 41, 73, and 84).

An often-quoted expression of Quetelet is quite meaningful, both in its appraisal of the history of science and in its message for us today. He stated: "The more advanced the sciences have become, the more they have tended to enter the domain of mathematics, which is a sort of center toward which they converge. We can judge the perfection to which a science has come by the facility more or less great, with which it may be approached by calculation" (Ref. [16], p. 39).

PROGRESS FOLLOWING QUETELET

Following Quetelet in the 19th Century there were many contributions to both development and application of the framework for analysis. This period served as the bridge between the creation of modern statistics by Quetelet and its rapid development in the 20th Century. The 20th Centruy also witnessed the development of computer science from key-punch and tabulation equipment through desk calculators to computers.

Florence Nightingale was greatly influenced by Quetelet, whom she admired as a scientific hero. Called the passionate statistician, she based her life upon the almost religious belief that administrators and legislators could be successful only if guided by the appropriate statistical data (Ref. [16], pp. 42 and 43).

In 1885, Hermann Ebbinghaus applied the normal distribution to the examination and analysis of higher mental processes. He was a pioneer in the application of precise quantitative methods to the analysis of data from purely psychological experiments. That work was founded upon Wilhelm Lexis' 1877 book on mass phenomena which clearly enunciated rules concerning the dispersion of statistical data and its relationship to the normality of the underlying distribution (Ref. [16], pp. 43 – 45).

One of the giants of this era was Francis Galton, who introduced both the percentile system and correlation. In 1874 and 1875, Galton proposed the percentile scale, including the idea of a median. His greatest contribution, the concept of correlation, was discovered prior to 1885. When Galton applied statistical analysis to anthropology, biology, and psychology, the diversity of topics amenable to statistical treatment was greatly increased (Ref. [16], pp. 46, 87, and 104).

Karl Pearson, another giant, made his original contributions during the same era. In 1893, Pearson invented the term, "moment," and the notation for moments about the mean and for those about an arbitrary origin, and he gave formulas for computing the former from the latter as well as describing a graphical device for computing moments. He initiated work on his forthcoming system of probability distributions in 1895 by stating relationships among the moments which would produce five types of distributions. Pearson and (Louis Napolean George) Filon in 1898 established the mathematical theory of – and general procedures for treating – two, three and four normally distributed random variables (Ref. [16], pp. 65, 73, and 74).

Jorgen Pedersen Gram in 1879 and Thorwald Nicolai Thiele in 1889 also investigated various types of probability distributions. In addition, the theory of regression and correlation was extended by Francis Ysidro Edgeworth in 1892, Karl Pearson in 1896, and George Udny Yule in 1897. Walter Frank Raphael Weldon applied correlation analysis to shrimp data in 1890 and to crab data in 1893 and William Lowe Bryan applied it to data on the motor ability of children in 1892 (Ref. [6], pp. 81 and 107–111).

ACKNOWLEDGMENTS

The author gratefully acknowledges the constructive comments of Francis J. Anscombe concerning clarification and expansion of the text. In addition, he is indebted to Victor Azgapetian, Wallace R. Blischke, and Leon Gainen for their critical review of the manuscript.

REFERENCES

[1] Bell, E. T., *Men of Mathematics* (Simon and Schuster, Inc., 1937).

[2] Boole, G., *Laws of Thought* (Dover Publications, Inc., reprint, original publication 1854).

[3] *Encyclopaedia Britannica* (Encyclopaedia Britannica, Inc., 1961).

[4] Goodman, A. F., "Flow of Scientific and Technical Information: The Results of a Recent Major Investigation," Western Division Paper 4516 (AD657–558), McDonnell Douglas Astronautics Company, Aug. 1967 and Revised Apr. 1970.

[5] Goodman, A. F., L. Gainen, and C. O. Beum, Jr., "Complete System Analysis: Quantitative System Analysis: Computer Simulation, and System Optimization," Western Division Paper 4431 (N69–16331), McDonnell Douglas Astronautics Company, Mar. 1967 and Revised Dec. 1969.

[6] Horn, J., "Ready-Read Card," Datamation (May 1965).

[7] Luce, R. D. and H. Raiffa, *Games and Decisions* (John Wiley and Sons, Inc., 1958).

[8] "Mathematics," *Life Science Library* (Time, Inc., 1963).

[9] Muller, M. E., "Computers as an Instrument for Data Analysis, Technometrics (May 1970).

[10] Newman, J. R., *The World of Mathematics* (Simon and Schuster, Inc., 1956).

[11] Rosen, S., "Electronic Computers: A Historical Survey," Computing Surveys, (Mar. 1969).

[12] Sage, D. N., "Information Systems: A Brief Look into History," Datamation (Nov. 1968).

[13] Schussel, G., "IBM versus REMRAND," Datamation (May 1965).

[14] Struik, D. J., *A Concise History of Mathematics* (Dover Publications, Inc., 1948).

[15] *The Quiet Revolution: Computers Come of Age* (American Federation of Information Processing Societies, 1965).

[16] Walker, H. M., *Studies in the History of Statistical Method* (The Williams and Wilkins Company, 1929).

STATISTICAL CONCEPTS IN COMPUTATIONAL MATHEMATICS*

M. L. Juncosa

The Rand Corporation

ABSTRACT

For a young science and even for most old ones, the intensity and variety of activities involved in computer science are unsurpassed. In an effort to provide the extremely varied training needed in the field, curriculum planners have tended to slight probability and statistics. Rarely has it been a requisite and only occasionally a desired elective.

However, not only is an adequate foundation in probability and statistics necessary for such external computer applications as tracking and other data reduction, Monte Carlo techniques, gaming, operations research, traffic analyses, etc., but it also plays an important role in internal applications to computer science and technology.

This effort in Rand's continuing study of computational mathematics and its applications offers a number of examples in various areas of computational mathematics and computing machine arithmetic—to say nothing of computer design, time sharing, and the like—to support the thesis that probability and statistics should be a requisite, not simply an elective or, even worse, ignored, at any institution offering a program in computer science. Furthermore, a potentially profitable source of research problems can be found in the interface between probability and statistics and computer science.

I. INTRODUCTION

Only a few years ago, questions of the existence of a discipline called computer science were prevalent. Traditionally, a science is primarily identifiable at its core by its unifying axiomatics, whereas a technology is mostly distinguishable by its methodology and its resultant applications. Because computer science has not been sharply identified by these criteria to be either science or technology, it is easy to see why some scientific traditionalists would have questioned its existence as a science.

Computer science clearly has developed as a mixed breed, a product of logicians, mathematicians, engineers, physicists, psychologists, linguists, and many other scientists, both "hard" and "soft." Because of this healthy miscegenation, computer science has not had clearly defined, uniformly accepted requisites in university curricula.

Figure 1, from the March 1968 *Communications to the ACM* [3], shows how varied an undergraduate curriculum in computer science can be. The June 1967 Stony Brook conference on graduate curricula and research in computer science [6] demonstrated the same thing. The variety results because, as a young science, computer science means different things to different people, especially curriculum planners.

Furthermore, the undergraduate or graduate curricula generally offer many subjects allied to computer science, such as, logic, algebra, and graph theory in mathematics, and computer design in

*This research is supported by the United States Air Force under Project Rand. The topic developed herein has also been presented in invited talks before the 127th Annual Joint Statistical Meetings of ASA, ENAR, WNAR and IMS on Dec. 29, 1967, Computer Science and Statistics Third Annual Symposium on the Interface, January 31, 1969, and on a panel at the 1969 Association for Computing Machinery National Conference Aug. 27, 1969.

THE CORE COURSES OF THE PROPOSED UNDERGRADUATE PROGRAM

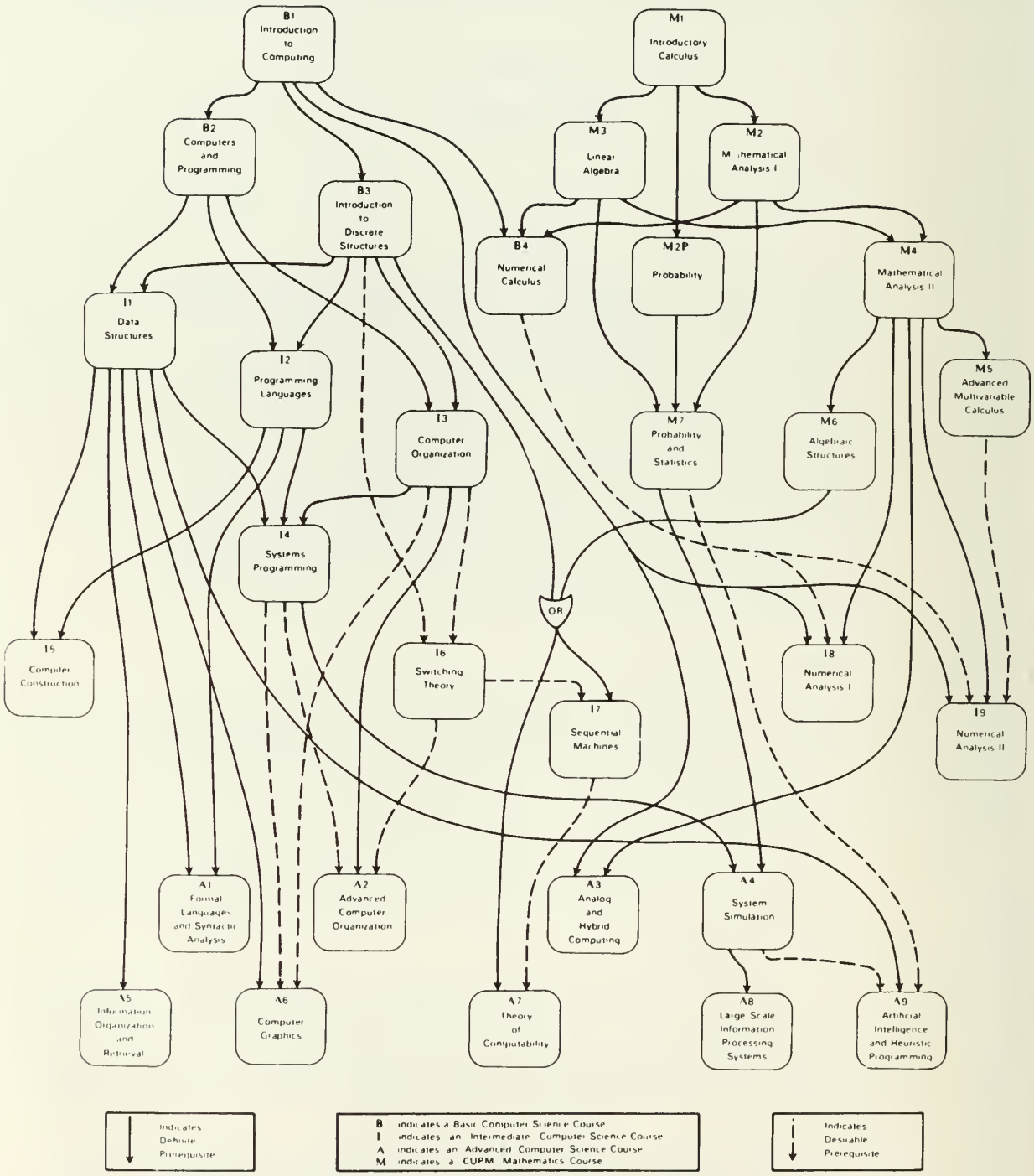

B1 Introduction to Computing
B2 Computers and Programming
B3 Introduction to Discrete Structures
M1 Introductory Calculus
M3 Linear Algebra
M2 Mathematical Analysis I
B4 Numerical Calculus
M2P Probability
M4 Mathematical Analysis II
I1 Data Structures
I2 Programming Languages
I3 Computer Organization
M5 Advanced Multivariable Calculus
M7 Probability and Statistics
M6 Algebraic Structures
I4 Systems Programming
OR
I5 Compiler Construction
I6 Switching Theory
I7 Sequential Machines
I8 Numerical Analysis I
I9 Numerical Analysis II
A1 Formal Languages and Syntactic Analysis
A2 Advanced Computer Organization
A3 Analog and Hybrid Computing
A4 System Simulation
A5 Information Organization and Retrieval
A6 Computer Graphics
A7 Theory of Computability
A8 Large Scale Information Processing Systems
A9 Artificial Intelligence and Heuristic Programming
Indicates Definite Prerequisite
B indicates a Basic Computer Science Course
I indicates an Intermediate Computer Science Course
A indicates an Advanced Computer Science Course
M indicates a CUPM Mathematics Course
Indicates Desirable Prerequisite

Figure 1. Prerequisite structure of courses.

engineering. However, only a few institutions offer probability and statistics as a requisite rather than an elective in their computer science curriculum. In those places where probability and statistics have been strongly recommended, one surmises that it is most often because they are considered good for applications, e.g., for operations research. Even Figure 1 seems to ignore their importance to computer science *per se*. And even among numerical analysts, probability and statistics are most often regarded as useful foundations for applications, e.g., to be able to use Monte Carlo methods to evaluate multi-dimensional integrals, to find extrema of functions of several variables, to solve ordinary and partial differential equations and integro-differential equations (like the transport equation), to do operations analyses, gaming, traffic simulation with queuing phenomena, etc., and to reduce data in a host of important applications.

It is regrettable that probability and statistics are not strongly required because in a number of diverse examples, they demonstrate a distinct pragmatic value directly to several problem areas of computer science, to say nothing of the applications of computer science to problems of probability and statistics.

In many areas of science, e.g., in physics, even when a problem has an inherent deterministic structure (particularly in some cases when many variables appear or a very complex process is involved), a statistical or probabilistic model may be devised to cut the Gordian knot. Such has been the case leading to the foundation of statistical mechanics.

In a classical gas, the molecular motions, in principle, are described for all time by their states at some given time and by the solution of the equations of Newtonian mechanics. But one may ask rhetorically, would one write equations expressing conservations laws for each and every one of some 10^{23} colliding particles. A reasonably satisfactory solution was found in the statistical approach of statistical mechanics. And, in atomic and nuclear physics going to smaller particles with still other fundamental difficulties, Newtonian mechanics gave way to probabilistic quantum mechanics.

We note, however, that no dogmatic ontological commitment is made as to whether or not, in the ultimate analysis, the driving processes at the core of the phenomena are random or not. It is primarily the pragmatism of obtaining manageable and useful results that gives rise to most statistical model making in any field. We shall see later that the study of accumulated errors in a lengthy calculation can be motivated by taking a similar statistical view.

Our main point in this work is to illustrate the existence of a certain indebtedness to probability and statistics for contributions to computer science. These contributions exist for concepts as well as practice, hardware, and software. Our main concern is with the contribution of probability and statistics to certain aspects of computational mathematics. This area is chosen because the author's interests are in applied mathematics with a considerable emphasis in computational mathematics.*

Although we do not detail the uses of probability and statistics in such areas of computer science as automata theory, programming, and hardware, a few applications are listed for them. W. H. Pierce [13] examines problems of reliability through redundancy in failure-tolerant computer design. Unsolved statistical problems, which he mentions, may stimulate a statistician to work on computer hardware problems. Allan L. Sherr [15] deals with traffic problems in timesharing and multi-programming. Both similar and other problems are discussed in Estrin and Kleinrock [5]. Estrin, Hopkins, Coggan, and

*By computational mathematics is meant any mathematics motivated by intentions to apply it to computation, computer design, etc. In most cases it is equivalent to numerical analysis, a poor terminology because it is not simply analysis of numbers. Another term is mathematics of computation.

Crocker [4] describe instrumentation of internal traffic-statistics gathering in a computer. With such knowledge and the application of some statistical theory, this approach should influence computer design. In the area of automata theory, Chaitin [2] uses some advanced theorems in probability theory to deal with the length of programs that compute finite binary sequences. However, we leave discussion of these topics, and others so intimately related to computer design and organization, to others working actively in the field.

Our general purpose here is to convince more computer science curriculum planners that probability and statistics *should* be a requirement rather than, at most, an elective. To this end, we support the thesis of substantial utility of probabilistic and statistical concepts in computer science *per se*, concentrating in the examples specifically on computational mathematics, and cite some areas for further research.

II. ELEMENTARY PROBABILITY AND ERRORS IN NUMERICAL ALGEBRAIC PROCESSES

Numerical treatment on a computer of such algebraic processes as solving linear algebraic equations and obtaining matrix eigenvalues and roots of polynomials nearly always approximates the desired result. We consider some pragmatic benefits derived from a statistical solving of deterministic problems.

Undoubtedly, once a program for a numerical problem is loaded, with its initial conditions set and computation started, the resulting process is purely deterministic. Each little round-off or chopping applied to the elementary arithmetic operations, and each truncation of a series, continued fraction, or other kind of approximation of a mathematical function or mathematical operation, contributes to the computation an error completely determined by the current state of the system and the input program. Furthermore, these errors frequently generate other errors that, like the loaded program, follow a determined path.

Unfortunately, this path is most tortuous and, in most cases, realistically incapable of explicit rendering. Consequently, even if only for lack of time, it is essentially impossible to determine the exact total error contribution to the answer at the end of the computation.

For example, the straightforward Gaussian elimination process for solving a system of n linear algebraic equations requires some $\frac{1}{3}\,n^3$ multiplications and $\frac{1}{3}\,n^3$ additions, each of which contribute some round-off or chopping error along the way. For $n = 100$, this results in some $\frac{2}{3}$ of a million unwanted error contributions.

Of course, in many problems one can overestimate each individual error, assume that it always goes in the same direction as the others, and add up the absolute values of all of the contributed errors as well as of all of the propagated errors, thus producing an *a priori* upper bound on the total error expected. Unfortunately, this upper bound most often is an inordinately pessimistic estimate of the error, particularly since it presumes the calamitous situation where each round-off or chopping error, or truncation error is not only as big as possible, but also always goes in the direction that makes the total error worse.

The use of such pessimistic estimates in planning a computation can frequently result in demanding double-precision computations, a procedure that quadruples computing time on machines, without double-precision hardware, if one is integrating differential equations or evaluating integrals, to take smaller steps which also increases the computing time and, in turn (for some, the most important item) the *bill* for computation.

A statistical approach to error analysis and other questions in numerical analysis provides an alternative by offering some realistic error estimates. The approach requires the user to adopt a philosophy of risk. When an undesired event of very low probability of occurrence results in a catastrophically negative payoff, one generally decides not to take the risk. That is, in a computation where imprecise results would be very costly, one would rely on pessimistic error bounds *guaranteed* not to be exceeded. On the other hand, if the costs were low, a gamble on nonoccurrence is reasonable. For such a policy, statistical estimates are a prerequisite.

In the case of inversion of high-order matrices, J. von Neumann and Herman Goldstine [8, 17], about 20 years ago, made the first big analysis of the accumulated errors in matrix computations and included some statistical estimates of them. J. H. Wilkinson in his 1961 paper [18], later summarized in his book [19], revisited the scene with a more detailed analysis, considering both fixed and floating point modes of computation as well as the use of single and double length accumulators. Although his rigorous error bounds are very detailed, his statistical estimates were given a much lighter treatment. A deeper statistical analysis, which desirably, would include correlations, would be a valuable contribution.

In essence, the statistical approach is based on a pragmatic model that assumes that round-off errors are independent and uniformly-distributed random variables. In such a sum of n rounded numbers, the expected error would be zero and the standard deviation of the error would be proportional to $\sqrt{n}$ times 2^{-s}, 10^{-s}, or 16^{-s}, if the machine word carried five places were binary, decimal, or hexadecimal, respectively, in its base for numerical-symbol representation. When the machine chops rather than rounds, and all summands are of the same sign, then the expected error grows linearly with n while the standard deviation is still, as expected, proportional to $\sqrt{n}$. On the other hand, absolute error bounds are proportional to n instead of $\sqrt{n}$.

The situation is more involved when products of numbers are also considered in the process. Elimination methods for solving linear algebraic systems involve products and sums, as do methods for polynomial evaluation and root-finding, and for such complicated situations as determination of matrix eigenvalues.

The following absolute error bound on the maximum residual in solving $Ax-b=0$ with s-bit floating point Gaussian elimination on a binary machine was given by Wilkinson [19]:

$$\max |Ax-b|_i < 2^{-s}g(2.005n^2+n^3+2^{-s-2}n^4) \max |x_i|,$$

where g is a bound on the elements of the successive matrices appearing in the triangularization process of the elimination. For reasonable s and the values of n manageable on current machines, the term $2^{-s-2}n^4$ is usually negligible. The dominant term is usually n^3; it results from the errors in the solution of the intermediate triangularized system which are proportional to n^2.

On the other hand, a direct statistical approach that pragmatically assumes independence of errors suggests dominant error-bound estimates proportional to $n^{3/2}$. However, Wilkinson notes that in his experience even errors of the order of n seem to be rarely exceeded. In the triangularization, whether partial or complete pivoting is used, it has been observed the growth of numbers appearing in the process very rarely approaches the bounds given for their growth. A statistical theory would be most welcome, particularly one with confidence limits.

Similar analyses have been tried on the simpler problem of fixed point evaluation of polynomials in a 1952 BRL report by Gorn and Juncosa [9] and in Wilkinson [19] and also on the harder problem of

eigenvalue determination. As in [9] consider the case of the fixed point evaluation of an nth order polynomial,

$$y_n = f(x) = a_0 x^{n-1} + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n,$$

with all numbers scaled in absolute value to be less than 1. Certain detailed modifications are needed for the floating point analysis, but the final results for the errors in the mantissa of y_n are essentially the same:

$$\epsilon(y_n) = f'(\bar{x})\epsilon(x) + \sum_{i=0}^{n-1} \epsilon(a_i) x^i + \beta^{-s} \sum_{i=0}^{n-1} \eta_i x^i,$$

where a bar over a symbol indicates the machine's evaluation of the number symbolized. Here, β is the machine number base, s is the word length (or mantissa length), and $\beta^{-s}\eta_i$ is the round-off error at the ith step of the polynomial-evaluation process by Horner's method. This round-off is complicated because it is given by a function hopeless to trace through each evaluation. A statistical approach with an assumption of a uniform distribution for η_i in $(-1/2, 1/2)$ yields

$$\sigma(\epsilon(y_n)) = \frac{\beta^{-s}}{\sqrt{12}} \sqrt{\frac{1-x^{2n}}{1-x^2}} \leq \frac{1}{2} \sqrt{\frac{n}{3}}\, \beta^{-s},$$

and at least p percent of the population of errors would fall within

$$\frac{p\sigma(\epsilon(y_n))}{100} \sqrt{\frac{3(1+x)(1-x^n)}{(1-x)(1+x^n)}},$$

of *the mean*.

In matrix eigenvalue problems, a general procedure is to reduce the matrix by orthogonal transformations to a Hessenberg matrix H, i.e., a matrix form with all zeros either below the diagonal or immediately below the main diagonal or its transpose. Wilkinson [19, p. 151] indicates that a statistical estimate of the difference between a true simple ith eigenvalue of an $n \times n$ matrix and its calculated value by the method of Morton Hyman [12] is proportional to

$$n \cdot 2^{-\beta} / |\cos \theta_i|.$$

where θ_i is the angle between the ith column of P and the ith row of P^{-1}, where the transformation, PHP^{-1}, reduces H to Jordan form.

III. STABILIZING NUMERICAL SOLUTIONS OF INTEGRAL EQUATIONS

The numerical solution of linear integral equations is frequently reduced to solving linear algebraic systems. Again, in the spirit of statistical mechanics, where the introduction of statistical considera-

tions led to a respectable, feasible treatment of insoluble deterministic problems, consider the problem of the solution for $y(t)$ of linear integral equations of the first kind,

$$\int_a^b K(s, t)y(t)dt = f(s),$$

with, for example, a continuous kernel, $K(s, t)$. The problem is ill-posed in the sense that $y(t)$ may not depend continuously on the data, $f(s)$, since a discontinuous $y(t)$ can yield a continuous $f(s)$. Small deviations from f can result in large deviations in y.

Solving the problem numerically is frequently done by discrete approximations to the integral equation which, if close, also tend to preserve the unstable character of the integral equation in the resulting algebraic system,

$$Ay = f.$$

Just as a juggler balancing a pole (a situation of unstable equilibrium) "dithers" the pole to give it a certain statistical stability, or a driver "dithers" his steering for the same purpose, recent approaches by Tihonov (regularization) [16] and Franklin (well-posed stochastic extension) [7] introduce random noise into the equation to induce numerical stability.

Following Franklin, the problem $Ay = f$ is replaced by

$$Au + n = g,$$

where A and g are known and u and n are samples chosen of random linear functionals, u_1 and u_2, that are elements of P_1 and P_2, random processes (i.e., collection of measurable linear functionals) on Hilbert spaces, H_1 and H_2, respectively. P_1 and P_2 are Borel families of linear functionals on H_1 and H_2 with associated probability measures. For u_1 and u_2 in P_1 and P_2, respectively, and from extraneous considerations (e.g., from *a priori* knowledge of computational noise, round-off, truncation, etc.), auto- and cross-correlation operators R_{11}, R_{22}, R_{12} are prescribed, i.e., $Eu_i u_j^* = R_{ij}$, where R_{ij} is a bilinear or a quadratic positive definite operator on $H_1 \times H_2$, and the asterisk denotes adjoint.

The processes, P_1 and P_2, define a third process, P_3, through

$$u_3 = Au_1 + u_2.$$

To "best" estimate the particular linear functional $u_1 = u$ in P_1 is, in Franklin's criterion, to estimate $u^* a_1$ for arbitrary a_1 in H with a given sample $u_3 = g$ from P_3. This is done by finding a_2 in H_2, such that

$$E|u_1^* a_1 - u_3^* a^2|^2 = \min,$$

which is

$$a_2 = R_{33}^{-1} R_{31} a_1.$$

Then, for arbitrary a_1, the best estimate for u_a^* is

$$g^* a_2 = g^*(R_{33}^{-1} R_{31} a_1) = (R_{13} R_{33}^{-1} g)^* a_1,$$

or the best estimate for u is

$$v = (R_{11}A^* + R_{12})(AR_{11}A^* + AR_{12} + R_{12}^* A^* + R_{22})^{-1} g.$$

If "signal" and "noise" processes, P_1 and P_2, are independent, this equation reduces to

$$v = R_{11}A^*(AR_{11}A^* + R_{22})^{-1}g.$$

With no noise and if R_{11} and A are nonsingular, we get, of course, the deterministic answer

$$v = A^{-1}g.$$

Prerequisite knowledge to understand such techniques includes familiarity with random processes, but the results prove worthwhile. The utility of the approach is the inversion of Poisson's integral formula,

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{(1-r^2)h(\phi)d\phi}{1-2r\cos(\theta-\phi)+r^2} = f(\theta).$$

For $r = 1/2$ we have

$$\frac{3}{2\pi}\int_0^{2\pi} \frac{h(\phi)d\phi}{5-4\cos(\theta-\phi)} = f(\theta).$$

Approximating this integral equation by the following linear algebraic one,

$$\frac{3}{N}\sum_{j=1}^{N} \frac{u(\theta_j)}{5-4\cos(\theta_i-\theta_j)} + n(\theta_i) = g(\theta_i),$$

with $N = 50$, Franklin shows startlingly good results. The true answer,

$$h(\phi) = \cos 3\phi - \cos\phi + \sin(\cos\phi)\cos(\sin\phi),$$

is less than 1 in absolute value. With no noise ($n(\theta_j) \equiv 0$) computational errors ran almost as bad as 2×10^{-2}. With appropriate random noise, errors ran around 5×10^{-6} or less.

Franklin assumed an autocorrelation function for $u_1(\theta)$ given by

$$Eu_1(\theta)u_1^*(\theta+\lambda) = \alpha e^{-\beta\sin^2\lambda/2},$$

the "size" of $u_1(\phi)$ given by

$$\sqrt{Eu^2(\theta)} = \sqrt{d},$$

and its "roughness" by

$$\sqrt{\frac{\overline{E[u_1'(\theta)]^2}}{Eu_1^2(\theta)}} = \sqrt{\frac{\beta}{2}}.$$

He computed with a "size" equal to unity and tried "roughnesses" 1, 2, 5, and 7 (the accuracies cited above obtained with roughness equal to 5), whereas the true "size" is given by

$$\sqrt{\frac{1}{50}\sum_{i=1}^{50} h^2(\theta_i)} = 0.59,$$

and the true "roughness" is 2.98.

IV. ERRORS IN NUMERICAL INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS

Returning to more conventionally trodden ground, in the area of integration of ordinary differential equations described as initial value problems, Dirk Brouwer [1], as early as 1937, remarked that a more realistic estimate of accumulated errors after n steps of numerical integration of a first order linear system would be proportional to $\sqrt{n}$, rather than n, the value given by rigorous conservative analysis of stable methods.

P. Henrici, in a more detailed analysis in 1963 [11, p. 42 *et seq.*] corroborated Brouwer's observations for linear ordinary differential equations and extended the theory. He also experimented with multistep methods (like Adams, Milne, and midpoint methods) on a linear and a nonlinear system of four equations. In the case of his nonlinear model, he found that rather than being proportional to $\sqrt{n}$ in the linear cases, the standard deviation of the error had an exponential growth. Furthermore, his experimental results confirmed the statistical theory.

Limitations of time and space prevent giving further details which are available in his book. However, a substantial challenge would be to attempt an extension of this type of analysis to the solution of initial value problems for partial differential equations.

V. SOME STATISTICAL CONSIDERATIONS IN COMPUTER ARITHMETIC PROCESSES

Turning to some more elementary remarks, we briefly mention some elementary statistical considerations on the efficiency of number bases used in a computer. The most common one has been the binary system. Initially, the decimal was internally employed on a few machines, e.g., the ENIAC. The octal system has also been employed, and currently the IBM 360 system has revived the use of the hexadecimal system. Furthermore, today almost all calculations are done in floating point, and normalizations of numbers (i.e., shiftings so that the leading figure in the mantissa or coefficient of the representation of the numbers is not zero) and generally performed, if necessary after most operations. The mantissa is frequently chosen to be less than one. An interesting question is the effect of number base on the frequency of normalization.

Pinkham [14] notes that well-used tables of logarithms are invariably dirtier in front than in back. This suggests that people are more interested in numbers with smaller leading digits than ones with larger ones. Some empirical statistics of large bodies of physical data support this. Pinkham references several papers and books dealing with the point, and he produces some mathematical basis for these statistics. R. W. Hamming also deals with this topic in his book on numerical analysis [10, p. 35 *et seq.*]. The choice of number base is clearly going to affect the frequency of normalization.

Suppose that numbers x and y, represented in a β-based machine, have mantissas ξ and η which are each independently and uniformly distributed (a moot assumption) in $(1/\beta, 1)$. This choice of interval guarantees that their initial figures are not zero. Consider their product. If the leading figure is zero, then a normalization shift causing loss of significant figure results. What is the probability of this event?

Since ξ and η are constrained to the interval $(1/\beta, 1)$, a normalization shift occurs when (ξ, η) is a point in the area A in the (ξ, η) plane (see Figure 2) which lies below the curve $\xi\eta = 1/\beta$, to the right of the line $\xi = 1/\beta$, and above the line $\eta = 1/\beta$. The probability of this event under the assumptions made is

$$\text{Prob}\left\{\text{shift in } \xi\eta\right\} = \frac{A}{(1-1/\beta)^2} = \frac{\beta \ln \beta - \beta + 1}{(\beta-1)^2},$$

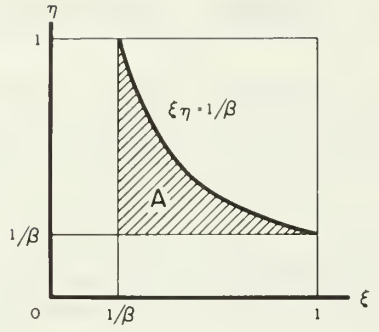


Figure 2. Normalization shift after multiplication.

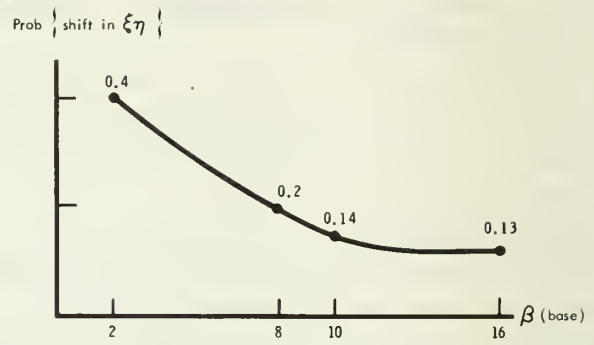


Figure 3. Probability of post-multiplication normalization shift.

and is plotted in Figure 3. For $\beta = 2, 8, 10$, and 16, we get the corresponding probabilities 0.4, 0.2, 0.14, and 0.13. The slope of the curve is

$$\frac{2+\ln\beta}{(\beta-1)^2}-\frac{2\beta\ln\beta}{(\beta-1)^3}=-\frac{(\beta+1)\ln\beta-2(\beta-1)}{(\beta-1)^3},$$

negative for all integers greater than 2.

Under the assumptions made, a hexadecimal machine is expected to make normalization shifts in floating point products about 1/3 as often as a binary machine. However, when this occurs there is about eight times the loss in significant figures. Thus, the factor of disadvantage in loss of accuracy is 8/3.

For division, say, ξ by η, the condition for post-division shift is that $\xi > \eta$ and the probability of shift is one-half, independent of β for machines not requiring a predivision shift. (Incidentally, sometimes a predivision shift is required in those computers which demand the quotient to be a fraction. Then no postdivision shift is needed.)

For addition of ξ and η, assumed to be positive numbers, the condition for a normalization shift after addition is $\xi+\eta > 1$. From geometrical considerations (see Figure 4), it is easily shown that the probability of a shift is

$$1-\frac{1}{2}\left(\frac{\beta-2}{\beta-1}\right)^2,$$

as depicted in Figure 5. This probability declines with increasing β to give for 2, 8, 10, and 16 as the probabilities of a postaddition shift of 1, 0.64, 0.604, and 0.56, respectively. With a hexadecimal machine, the factor of disadvantage over a binary one in addition of numbers of like sign is, therefore, around 9/2.

Unfortunately, this is neither the whole, nor even a precise story. In the floating point representation of the original numbers x and y, their respective exponents are quite likely to be different. Thus, a preaddition denormalization shift on the smaller number is necessary to make the exponents equal in order to start the computation of a sum in floating point. We need a more detailed analysis, or at least empirical evidence, of the distribution of exponents. Although one is lacking, it is clear that the frequency of shifts goes down rapidly as β increases.

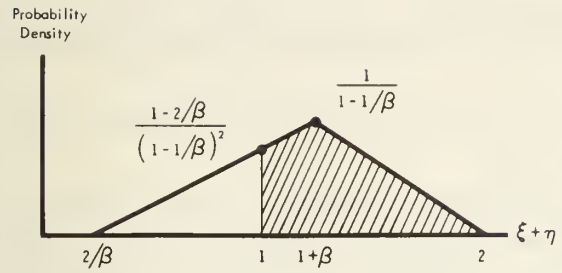


Figure 4. Normalization shift after addition.

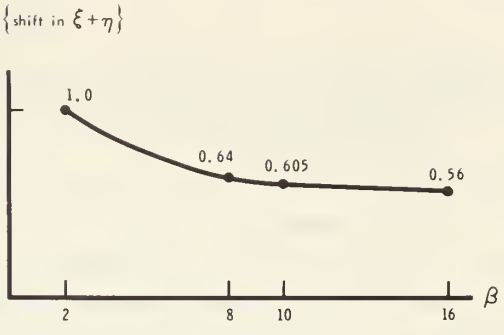


Figure 5. Probability of post-addition normalization shift.

Note also that these analyses are made with pristine numbers, assumed to satisfy initially a uniform distribution. However, as noted earlier, there is a general tendency toward logarithmic distributions of initial digits (or whatever name one chooses to give to single-word numerals in bases other than the decimal). Therefore, as these numbers undergo some processing, the cited factors of disadvantage for later operations may not be as large. A more tedious exercise is to obtain appropriate formulas for probabilities that correspond to those given for numbers that were originally uniformly distributed random variables.

On a more difficult level, we need a statistical theory for shift frequencies in longer calculations which are composed of a mixture of arithmetic processes. One fly in the ointment is the nonassociative character of actual digital computer processing of numerical calculations which relies on finite word lengths. That is, the order in which the computation is executed affects the error obtained in the final result. However, in a first attempt, one may be entitled to ignore this, particularly in a statistical treatment.

Nonetheless, from the simple remarks made in this section, one sees that, if the prerequisites of greater familiarity with more advanced computational mathematics are too demanding, there is still plenty of room for significant statistical analysis at the level of simple machine arithmetic operations.

VI. CONCLUSION

Several examples have demonstrated the dependence of computer science, particularly computational mathematics, upon results from probability and statistics. We also cite a few problems and avenues for further exploration. Although we give short shrift to applications of probability and statistics in other areas of computer science, they are considerably more numerous than in computational mathematics, particularly in the areas of time sharing, multiprogramming, and traffic flow through a computer.

Nevertheless, by citing examples and potential areas for further research, even though it is only in the one area of computational mathematics, we hope to stimulate those who influence computer science curricula to replace occasionally encountered electives of probability and statistics by their requirement. Furthermore, those seeking problem areas for graduate research could do well to examine the interface between probability and statistics and computer science.

REFERENCES

[1] Brouwer, D., "On the Accumulation of Errors in Numerical Integration," *Astronomical Journal 46*, 149–153 (1937).

[2] Chaitin, Gregory J., "On the Length of Programs for Computing Finite Binary Sequences: Statistical Considerations," *J. ACM 16*, 145–159 (Jan. 1969).

[3] "Curriculum 68, Recommendations for Academic Programs in Computer Science," Report of the ACM Curriculum Committee on Computer Science, *Commun. ACM 11*, 151–197 (Mar. 1968).

[4] Estrin, G., D. Hopkins, B. Coggan, and S. D. Crocker, "SNUPER COMPUTER—A Computer in Instrumentation Automaton," *Proc. SJCC* (1967), pp. 645–656.

[5] Estrin, G. and L. Kleinrock, "Measures, Models and Measurements for Time-Shared Computer Utilities," *Proc. ACM* (1967), pp. 85–96.

[6] Finerman, Aaron (ed.), *University Education in Computing Science* (Academic Press, New York, 1968).

[7] Franklin, Joel N., "Well-Posed Stochastic Extensions of Ill-Posed Linear Problems," *Programming Report No. 135*, California Institute of Technology, Willis H. Booth Computing Center, Aug. 1969. Also *J. Math. Anal. and Appl., 31*, 3, 682–716 (1970).

[8] Goldstine, H. H. and J. von Neumann, "Numerical Inverting of Matrices of High Order," *Proc. Am. Math. Soc. 2*, 188–202 (Apr. 1951).

[9] Gorn, S. and M. L. Juncosa, "Some Estimates of Confidence Limits of Errors in Evaluating Polynomials and Adding Numbers," *Memorandum Report No. 624*, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland, 1952.

[10] Hamming, R. W., *Numerical Methods for Scientists and Engineers* (McGraw Hill Book Company, Inc., New York, 1962).

[11] Henrici, Peter, *Error Propagation for Difference Methods* (John Wiley and Sons, Inc., New York, 1963).

[12] Hyman, Morton A., "Eigenvalues and Eigenvectors of General Matrices," Univac Applications Research Center, *Technical Report No. 13*, July 1957 (presented to the American Mathematical Society (Sept. 1956; Apr. 1957) and to the ACM (June 1957)).

[13] Pierce, William H., *Failure-Tolerant Computer Design* (Academic Press, New York, 1965).

[14] Pinkham, Roger S., "On the Distribution of First Significant Digits," *Ann. Math. Stat. 32*, 1223–1230 (Dec. 1961).

[15] Scherr, Allan L., "An Analysis of Time-Shared Computer Systems," *Research Monograph No. 36* (The M.I.T. Press, Cambridge, Massachusetts, 1967).

[16] Tihonov, A. N., *Solution of Incorrectly Formulated Problems and the Regularization Method*, Dokl. Akad. Nauk SSR 151 (1963), pp. 501–504 (in Russian); Soviet Math. Dokl. 4 (1963) p. 1035 (English translation).

[17] von Neumann, John and H. H. Goldstine, "Numerical Inverting of Matrices of High Order," *Bull. Am. Math. Soc. 53*, 1021–1099 (Nov. 1947).

[18] Wilkinson, J. H., "Error Analysis of Direct Methods of Matrix Inversion," *J. ACM, 8*, 281–330 (July 1961).

[19] Wilkinson, J. H., *Rounding Errors in Algebraic Processes* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963).

EXTENDED ITERATIVE WEIGHTED LEAST SQUARES: ESTIMATION OF A LINEAR MODEL IN THE PRESENCE OF COMPLICATIONS

A. F. Goodman

McDonnell Douglas Astronautics Company—Western Division
Huntington Beach, California

ABSTRACT

This paper introduces an extended iterative weighted least squares procedure, denoted by EIWLS, for solution of a classical problem in analysis of scientific and technical information: estimation in the presence of complications, of the coefficients for linearly independent component signals in a linear model, from observations on the component signals and a composite signal containing the linear model plus noise which is nonstationary and/or correlated with unknown covariance matrix. An iterative weighted least squares procedure, denoted by IWLS, is developed for estimation in the absence of complications. Then IWLS is extended to perform the estimation subject to: (1) estimators being consistent with a priori information describing the random variation of coefficients over all possible states of nature (e.g., all systems of a specified type from a production process); (2) utilization of data from all pertinent channels in the estimation of coefficients which appear in the linear models for more than one data channel; and (3) replacment of the linear model by a new linear model containing only representative component signals which are highly-descriptive, but not highly-related, when there is a large number of component signals in the linear model and some of them are highly-related. FORTRAN computer programs have been written to implement IWLS and EIWLS on the IBM 7094 for the case of nonstationary and uncorrelated noise.

1. INTRODUCTION AND SUMMARY

Since the early 19th century, estimation of a linear model from data subject to error has been a classical problem in analysis of scientific and technical information. The least squares procedure for solution of the problem, formulated by Gauss in 1802 and published by Legendre in 1806, is essentially the first statistical technique developed for analysis of information [13]. However, the effectiveness of least squares and related procedures mainly depends upon characteristics of the error.

This paper introduces an extended iterative weighted least squares procedure, denoted by EIWLS, for solution of the classical problem in the presence of complications. A complete description of the problem, which may be termed a generalized statistical regression problem, is as follows: estimation in the presence of complications, of the coefficients for linearly independent component signals in a linear model, from observations on the component signals and a composite signal containing the linear model plus noise with unknown characteristics. Component signals are sometimes called input or independent variables, the composite signal is sometimes called an output or dependent variable, and the noise is sometimes called a residual or random error. Since EIWLS was developed originally for error analysis of an inertial navigation system, the signal-and-noise terminology is employed.

Pertinent characteristics of the noise are contained in the square array of noise variances and covariances, called the noise covariance matrix. Noise is said to be stationary and uncorrelated when that matrix is a constant multiple of the identity matrix, nonstationary and uncorrelated when it is a

diagonal matrix, stationary and correlated when each row of it is a proper arrangement of elements in the first row, and nonstationary and correlated otherwise.

Consider estimation in the absence of complications. If the noise covariance matrix is known, then the optimum estimators of the coefficients are the weighted least squares estimators determined by its inverse. If the noise covariance matrix is not known, its estimation appears to be a reasonable step toward estimation of the coefficients.

Estimation of the linear model is, itself, required to estimate the noise covariance matrix. Goodman [9] presented an iterative weighted least squares procedure, denoted by IWLS, to accomplish the estimation when the noise is nonstationary and/or correlated (i.e., not stationary and uncorrelated—in which case, the least squares estimators are optimum) with unknown covariance matrix. Briefly, IWLS:

1. Obtains the least squares estimators of the coefficients.

2. Calculates an estimator of the noise covariance matrix by using the composite signal and its least squares estimator, based upon the least squares estimators of the coefficients, to estimate the necessary noise variances and covariances.

3. Obtains the weighted least squares estimators of the coefficients which are determined by the inverse of this matrix estimator.

4. Iteratively repeats 2 and 3, with the least squares estimators of the coefficients replaced by the latest set of weighted least squares estimators of the coefficients, and obtains a new estimator of the noise covariance matrix and a new set of weighted least squares estimators of the coefficients.

5. Continues the iteration in 4 until a preassigned level of stability is attained.

The complications and their treatment were also summarized in general terms by Goodman [9]. This paper is an extension of Ref. [9] and introduces an improved and extended IWLS. Improvement of IWLS, as presented in Ref. [9], involves improved estimation of the noise covariance matrix. In the following paragraphs, each complication and the corresponding extension of IWLS to EIWLS is briefly discussed.

Coefficients in the linear model are constant for a particular state of nature (e.g., a particular system of a specified type from a production process). However, they may vary randomly from one state of nature to another. A priori information describing the coefficients' random variation, over all possible states of nature (e.g., all systems of that specified type from the production process), may exist from previous analysis; and the estimators ought to be consistent with it. To insure this, a modification of IWLS permits the incorporation into the procedure of a priori information concerning the means and covariance matrix of the coefficients.

Data may exist from several data channels, and a coefficient may appear in the linear models for more than one data channel. The data from all pertinent channels should be utilized in the estimation of that coefficient. This may be accomplished by properly arranging the coefficients and data from all channels into a form suitable for the application of IWLS.

There may be a large number of component signals in the linear model and some of them, though linearly independent, may be highly-related. Component signals are called highly-related in this paper when they possess a high degree of linear dependence. For accuracy and ease of computation, it is frequently desirable to replace the linear model by a new linear model containing only representative component signals which are highly-descriptive, but not highly-related, and to estimate the coefficients of the representative component signals in the new linear model. To accomplish this, the set of component signals is partitioned into subsets of highly-related ones and the appropriate weighted

average of a subset is selected to be the representative component signal for that subset in the new linear model. The coefficients in the new linear model may then be estimated by IWLS. In addition, an estimator for the coefficient of the representative component signal for a subset is apportioned among the coefficients of component signals in the subset, via the weighting scheme used in the selection of that respresentative component signal.

Additional operations, such as pre-editing of data and inclusion of estimates based upon previous sets of data into the a priori information, may be added without too much difficulty.

Although EIWLS is applicable when the noise is nonstationary and/or correlated, it has been programmed for the computer only in the case of nonstationary and uncorrelated noise. Two FORTRAN computer programs, the basic one, Ref. [18], and another, Ref. [17], using Efroymson's technique, Ref. [8], to preselect the representative component signals, have been written to implement IWLS as presented in Ref. [9]; and a FORTRAN computer program, Ref. [5], has been written to implement EIWLS.

Iterative statistical procedures such as EIWLS extract considerably more information from the data than do noniterative, closed-form procedures such as least squares. In view of recent computer hardware and software development, iterative procedures have been feasible to implement and evaluate for some time and continue to become more so with the passage of time. It is therefore "penny-wise and pound-foolish" not to utilize EIWLS, when dictated by theory, statistical tests such as the one given in Ref. [31], or examination of the data. Indeed, EIWLS is even more appropriate today than at the time of its development.

Since iterative statistical procedures bridge the gap between the two extremes of noniterative, closed-form and optimum statistical procedures, it is somewhat surprising that the development of a meaningful theory for iterative statistics has been essentially neglected in favor of the continued—and almost academic—characterization and comparison of the two extremes. This is well illustrated by the survey of related literature in section 7. It is noteworthy that the recent Ref. [31] contains an iterative procedure which is closely-related to IWLS, as well as an approach to the construction of confidence intervals and statistical tests, when the noise is nonstationary and uncorrelated.

Reference [13] recommends the consideration of four questions regarding an iterative statistical procedure:

1. Under what conditions does the iterative procedure converge?
2. How rapidly does the iterative procedure converge?
3. Under what conditions does the iterative procedure converge to the proper solution?
4. To what extent does the iterative procedure improve upon a noniterative, closed-form procedure?

Partial answers to all four questions are provided in sections 2–8.

Those interested in only the essence of EIWLS may confine themselves to section 1. Sections 6, 7, and 8 augment section 1 with a discussion, a comprehensive survey of related literature, and an illustrative example. In sections 2 through 5, the statistical details of EIWLS (whose comprehension may require careful reading by the nonstatistician) are displayed, with a minimum of development and amplification, along with some reasonable alternatives which are listed in footnotes.

2. ITERATIVE WEIGHTED LEAST SQUARES

Suppose that t represents time or some other auxiliary variable. At time t: $X_1(t)$, $X_2(t)$, . . ., $X_p(t)$ denote linearly independent component signals; β_1, β_2, . . ., β_p denote unknown coefficients;

$Y(t)$ denotes a composite signal; and $\epsilon(t)$ denotes noise with mean zero, variance $\sigma^2(t)$ and covariance $\sigma(t, t')$ between $\epsilon(t)$ and $\epsilon(t')$. Then the linear model is $\sum_{j=1}^{p} \beta_j X_j(t)$ and the representation of $Y(t)$ as containing the linear model plus noise is

$$Y(t) = \sum_{j=1}^{p} \beta_j X_j(t) + \epsilon(t). \tag{1}$$

Given the explicit observations X_{ji} of $X_j(t)$ for $j = 1, 2 \ldots, p$ and Y_i of $Y(t)$, which imply the implicit observation $\epsilon_i = Y_i - \sum_{j=1}^{p} \beta_j X_{ji}$ of $\epsilon(t)$, at time t_i for $i = 1, 2, \ldots, n > p$, the generalized statistical regression problem without the complicating restrictions is the estimation of $\beta_1, \beta_2, \ldots, \beta_p$ when $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ have mean zero and unknown n-by-n covariance matrix Σ of variances $\sigma_{ii} = \sigma_i^2 = \sigma^2(t_i)$ and covariances $\sigma_{hi} = \sigma(t_h, t_i)$.

Complicated expressions in this and subsequent paragraphs may be written in compact form by the introduction of matrix notation. Let

$$\underline{X}_j = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \cdot \\ \cdot \\ \cdot \\ X_{jn} \end{bmatrix} \text{ for } j = 1, 2, \ldots, p,$$

$$X = (\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p),$$

$$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix},$$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix},$$

$$\underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix} \quad \text{and}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{bmatrix}$$

The matrix alter ego of Eq. (1) is

$$\underline{Y} = \sum_{j=1}^{p} \beta_j \underline{X}_j + \underline{\epsilon} = X\underline{\beta} + \underline{\epsilon}; \tag{2}$$

and the matrix alter ego of the generalized statistical regression problem without the complicating restrictions is the estimation of $\underline{\beta}$ given X and $\underline{Y}$, when Σ is unknown.

The weighted least squares estimators of $\beta_1, \beta_2, \ldots, \beta_p$ which are determined by the n-by-n matrix

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{12} & w_{22} & \cdots & w_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ w_{1n} & w_{2n} & \cdots & w_{nn} \end{bmatrix}$$

of weights w_{hi} (or any matrix that is a constant multiple of W) are those minimizing the quadratic form

$$\begin{aligned} Q_W &= \sum_{h=1}^{n} \sum_{i=1}^{n} w_{hi} (Y_h - \sum_{j=1}^{p} \beta_j X_{jh}) \Big(Y_i - \sum_{j=1}^{p} \beta_j X_{ji} \Big) \\ &= (\underline{Y} - X\underline{\beta})' W (\underline{Y} - X\underline{\beta}). \end{aligned}$$

It may be shown that these estimators are given by

$$\hat{\underline{\beta}}_W = (\hat{\beta}_{W1}, \hat{\beta}_{W2}, \ldots, \hat{\beta}_{Wp})' = (X'WX)^{-1} X'W\underline{Y}$$

and have covariance matrix $(X'WX)^{-1}X'W\Sigma WX(X'WX)^{-1}$. They are called the least squares estimators if W is the n-by-n identity matrix I_n.

In the sense of possessing minimum variance among all unbiased linear estimators, of providing an estimator of $\sum_{j=1}^{p} \beta_j X_{ji}$ which has minimum variance among all unbiased linear estimators of it and of being maximum likelihood estimators when the noise is normally distributed, the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ are the weighted least squares estimators which are determined by Σ^{-1} (Ref. [2], chap. 14 and Ref. [32], sect. 1.3–1.5). The least squares estimators are, therefore, optimum only if the noise is stationary and uncorrelated, and Σ is a constant multiple of the identity matrix. For nonstationary and/or correlated noise, it is not unreasonable to presume that estimation of Σ and then estimation of $\beta_1, \beta_2, \ldots, \beta_p$ by weighted least squares is superior to estimation of $\beta_1, \beta_2, \ldots, \beta_p$ by least squares. Estimation of $\sum_{j=1}^{p} \beta_j X_{ji}$ for $i = 1, 2, \ldots, n$ is, however, required to estimate Σ. This may be accomplished via estimation of $\beta_1, \beta_2, \ldots, \beta_p$. Hence, iterative estimation of $\beta_1, \beta_2, \ldots, \beta_p$ and then Σ, as accomplished by IWLS should provide improvement over least squares estimation of $\beta_1, \beta_2, \ldots, \beta_p$.

Stated in symbols, IWLS:

1. Obtains the least squares estimators $\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \ldots, \hat{\beta}_p^{(0)}$ from

$$\underline{\hat{\beta}}^{(0)} = (\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \ldots, \hat{\beta}_p^{(0)})' = (X'X)^{-1}X'\underline{Y}. \tag{3}$$

2. Calculates an estimator $\hat{\Sigma}^{(1)}$, by using Y_i and its least squares estimator

$$\hat{Y}_i^{(0)} = \sum_{j=1}^{p} \hat{\beta}_j^{(0)} X_{ji}. \tag{4}$$

for $i = 1, 2, \ldots, n$ in an appropriate estimation scheme to calculate the necessary $\hat{\sigma}_i^{(1)2}$ and $\hat{\sigma}_{hi}^{(1)}$.

3. Obtains the weighted least squares estimators $\hat{\beta}_1^{(1)}, \hat{\beta}_2^{(1)}, \ldots, \hat{\beta}_p^{(1)}$ which are determined by $(\hat{\Sigma}^{(1)})^{-1}$ from

$$\begin{aligned} \underline{\hat{\beta}}^{(1)} &= (\hat{\beta}_1^{(1)}, \hat{\beta}_2^{(1)}, \ldots, \hat{\beta}_p^{(1)})' \\ &= [X'(\hat{\Sigma}^{(1)})^{-1}X]^{-1}X'(\hat{\Sigma}^{(1)})^{-1}\underline{Y}. \end{aligned} \tag{5}$$

4. Iteratively repeats 2 and 3, with $\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \ldots, \hat{\beta}_p^{(0)}$ and $\hat{Y}_i^{(0)}$ for $i = 1, 2, \ldots, n$ replaced by the latest set of weighted least squares estimators $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ and

$$\hat{Y}_i^{(c)} = \sum_{j=1}^{p} \hat{\beta}_j^{(c)} X_{ji} \tag{6}$$

for $i = 1, 2, \ldots, n$, and obtains a new estimator $\hat{\Sigma}^{(c+1)}$ and a new set of weighted least squares estimators $\hat{\beta}_1^{(c+1)}, \hat{\beta}_2^{(c+1)}, \ldots, \hat{\beta}_p^{(c+1)}$ from

$$\begin{aligned} \underline{\hat{\beta}}^{(c+1)} &= (\hat{\beta}_1^{(c+1)}, \hat{\beta}_2^{(c+1)}, \ldots, \hat{\beta}_p^{(c+1)})' \\ &= [X'(\hat{\Sigma}^{(c+1)})^{-1}X]^{-1}X'(\hat{\Sigma}^{(c+1)})^{-1}\underline{Y}. \end{aligned} \tag{7}$$

5. Continues the iteration in 4 for $c = 1, 2, \ldots, c^* - 1$, where c^* either is determined when a valid measure of the change in $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ becomes less than a preassigned constant or is, itself, a preassigned constant. An estimator for the covariance matrix of $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ is given by $[X'(\hat{\Sigma}^{(c)})^{-1}X]^{-1}X'(\hat{\Sigma}^{(c)})^{-1}\hat{\Sigma}^{(c+1)}(\hat{\Sigma}^{(c)})^{-1}X[X'(\hat{\Sigma}^{(c)})^{-1}X]$, which simplifies to $[X'(\hat{\Sigma}^{(c)})^{-1}X]$ when $\hat{\Sigma}^{(c+1)} = \hat{\Sigma}^{(c)}$, for $\hat{\Sigma}^{(0)} = I_n$ and $c = 0, 1, \ldots, c^*$.

The selection of an appropriate estimation scheme to use in 2 is influenced not only by the type of noise and the data, but also by statistical considerations.

When the noise is nonstationary and uncorrelated,

$$\Sigma = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2).$$

An estimator of σ_i^2 is provided by

$$\hat{\sigma}_i^{(c)2} = (Y_i - \hat{Y}_i^{(c-1)})^2 \text{ for } c = 1, 2, \ldots, c^* \text{ and } i = 1, 2, \ldots, n. \tag{8}$$

These estimators are somewhat unsatisfactory because each of them is based upon only one observation; and they should be combined to produce more satisfactory estimators. In most applications, it is reasonable to assume a linear model for the variation of $\sigma^2(t)$. Let $T_1(t), T_2(t), \ldots, T_P(t)$ be known functions of t and $\nu_1, \nu_2, \ldots, \nu_P$ be unknown coefficients with $p + P < n$; and assume

$$\sigma^2(t) = \sum_{k=1}^{P} \nu_k T_k(t).$$

A simple version of $\sum_{k=1}^{P} \nu_k T_k(t)$, which is employed by the computer program, Ref. [5], is the polynomial $\sum_{k=1}^{P} \nu_k t^{k-1}$. Suppose that $T_{ki}=T_k(t_i)$ for $i=1, 2, \ldots, n$ and $k=1, 2, \ldots, P$,

$$\underline{T}_k = \begin{bmatrix} T_{k1} \\ T_{k2} \\ \\ \cdot \\ \cdot \\ \cdot \\ T_{kn} \end{bmatrix} \text{ for } k=1, 2, \ldots, P,$$

$$T=(\underline{T}_1, T_2, \ldots, T_P),$$

$$\underline{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \\ \cdot \\ \cdot \\ \cdot \\ \nu_P \end{bmatrix} \text{ and}$$

$$\underline{\hat{\sigma}}^{(c)2} = \begin{bmatrix} \hat{\sigma}_1^{(c)2} \\ \hat{\sigma}_2^{(c)2} \\ \\ \cdot \\ \cdot \\ \cdot \\ \hat{\sigma}_n^{(c)2} \end{bmatrix} \text{ for } c=1, 2, \ldots, c^*.$$

Then an appropriate estimation scheme to use in 2 is

$$\underline{\hat{\nu}}^{(1)}=(\hat{\nu}_1^{(1)}, \hat{\nu}_2^{(1)}, \ldots, \hat{\nu}_P^{(1)})', =(T'T)^{-1}T'\underline{\hat{\sigma}}^{(1)2}, \tag{9}$$

$$\underline{\hat{\nu}}^{(c)}=(\hat{\nu}_1^{(c)}, \hat{\nu}_2^{(c)}, \ldots, \hat{\nu}_p^{(c)})'$$

$$=\{T'[(\hat{\Sigma}^{(c-1)})^{-1}]^2T\}^{-1}T'[(\hat{\Sigma}^{(c-1)})^{-1}]^2\hat{\sigma}^{(c)2} \tag{10}$$

for $c=2, 3, \ldots, c^*$ and

$$\hat{\Sigma}^{(c)}=\operatorname{diag}\left(\sum_{k=1}^{P}\hat{\nu}_k^{(c)}T_{k1}, \sum_{k=1}^{P}\hat{\nu}_k^{(c)}T_{k2}, \ldots, \sum_{k=1}^{P}\hat{\nu}_k^{(c)}T_{kn}\right) \tag{11}$$

for $c=1, 2, \ldots, c^*$. It might be observed that the use of Eqs. (9)–(11) produces a simple iterative procedure for obtaining a solution to the maximum likelihood equations whose accuracy may be checked for nonstationary, uncorrelated noise which is normally distributed with $\sigma^2(t)=\sum_{k=1}^{P}\nu_k T_k(t)$. To establish an upper limit for all weights, all estimators $\sum_{k=1}^{P}\hat{\nu}_k^{(c)}T_{ki}$ of σ_i^2 must be bounded away from zero by a lower limit.

Consider the case of stationary, correlated noise, and let the observations be taken at different and equally-spaced times with $t_i = t_0 + i\Delta t$. Then $\sigma(t, t') = \sigma(\Delta)$ is a function of only the separation time $\Delta = |t - t'|$, $\sigma_{hi} = \sigma(\tau)$ is a function of only $\tau = |h - i|$ and

$$\Sigma = \begin{bmatrix} \sigma(0) & \sigma(1) & \sigma(2) & \cdots & \sigma(n-1) \\ \sigma(1) & \sigma(0) & \sigma(1) & \cdots & \sigma(n-2) \\ \sigma(2) & \sigma(1) & \sigma(0) & \cdots & \sigma(n-3) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \sigma(n-1) & \sigma(n-2) & \sigma(n-3) & \cdots & \sigma(0) \end{bmatrix}.$$

An estimator ‡ of $\sigma(\tau)$ is

$$\hat{\sigma}^{(c)}(\tau) = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} (Y_i - \hat{Y}_i^{(c-1)})(Y_{i+\tau} - \hat{Y}_{i+\tau}^{(c-1)}) \tag{12}$$

for $c = 1, 2, \ldots, c^*$ and $\tau = 0, 1, \ldots, n-1$.

It is usually reasonable to assume a linear model for the variation of $\sigma(\Delta)$. Suppose that $T_1(\Delta)$, $T_2(\Delta)$, $\ldots$, $T_P(\Delta)$ are known functions of Δ, $\nu_1, \nu_2, \ldots, \nu_P$ are unknown coefficients with $p + P < n$ and

$$\sigma(\Delta) = \sum_{k=1}^{P} \nu_k T_k(\Delta),$$

whose simple version is

$$\sigma(\Delta) = \sum_{k=1}^{P} \nu_k \Delta^{k-1}.$$

If $T_{ki} = T_k((i-1)\Delta t)$ for $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, P$,

$$\underline{T}_k = \begin{bmatrix} T_{k1} \\ T_{k2} \\ \cdot \\ \cdot \\ \cdot \\ T_{kn} \end{bmatrix} \quad \text{for } k = 1, 2, \ldots, P,$$

$$T = (\underline{T}_1, \underline{T}_2, \ldots, \underline{T}_P),$$

$$\underline{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \cdot \\ \cdot \\ \cdot \\ \nu_P \end{bmatrix},$$

‡An alternative estimator, which takes the estimation of $\beta_1, \beta_2, \ldots, \beta_p$ explicitly into account, divides the sum by $n-p-\tau$ and limits τ to $\tau = 0, 1, \ldots, n-p-1$.

$$\underline{\hat{\sigma}}^{(c)} = \begin{bmatrix} \hat{\sigma}^{(c)}(0) \\ \hat{\sigma}^{(c)}(1) \\ \\ \vdots \\ \\ \hat{\sigma}^{(c)}(n-1) \end{bmatrix} \quad \text{and}$$

$$N = \text{diag}\,(n,\ n-1,\ \ldots,\ 1),$$

then an appropriate estimation scheme to use in 2 is given by

$$\underline{\hat{\nu}}^{(c)} = (\hat{\nu}_1^{(c)},\ \hat{\nu}_2^{(c)},\ \ldots,\ \hat{\nu}_P^{(c)})' = (T'NT)^{-1}T'N\underline{\hat{\sigma}}^{(c)} \tag{13}$$

and

$$\sum{}^{(c)} = \begin{bmatrix} \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k3} \ \ldots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{kn} \\ \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} \ \ldots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,\,n-1} \\ \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k3} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} \ \ldots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,\,n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{kn} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,\,n-1} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,\,n-2} \ \ldots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} \end{bmatrix} \tag{14}$$

for $c = 1, 2, \ldots, c^*$.

The weight matrix N is used in Eq. (13) because it constitutes a simple means of reflecting the increase of statistical dependability in $\hat{\sigma}^{(c)}(\tau)$ with $n-\tau$. It is necessary to require that all estimators $\left|\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{ki}\right|$ of $\sigma^{(c)}(\tau)$ be bounded above by the estimator $\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1}$ of $\sigma^{(c)}(0)$. A common engineering model for the variation of $\sigma(\Delta)$ is

$$\sigma(\Delta) = \sigma_0 e^{-\nu\Delta},$$

which is nonlinear in the unknown coefficients σ_0 and ν. This exponential variation of $\sigma(\Delta)$ may be treated† by transforming the model to the linear one,

$$\log_e \sigma(\Delta) = \log_e \sigma_0 - \nu\Delta = \nu_0 - \nu\Delta,$$

and proceeding as above with $\hat{\sigma}^{(c)}(\tau)$, T_{1i}, T_{2i}, ν_1, and ν_2 being replaced by $\log_e \hat{\sigma}^{(c)}(\tau)$, 1, $(i-1)\Delta t$, ν_0 and $-\nu$, respectively.

---

†If preferred, $\sigma_0 e^{-\nu\Delta}$ may be approximated by $\sum_{k=1}^{P'} \nu_k \Delta^{k-1}$ and $\nu_1, \nu_2, \ldots, \nu_P$ may be estimated as in Eq. (13); or σ_0 and ν may be estimated by those (nonlinear) estimators that minimize $\sum_{i=1}^{n} (n-i+1)[\hat{\sigma}^{(c)}(i-1) - \sigma_0 e^{-\nu(i-1)\Delta t}]^2$, but are difficult to compute.

A proper modification of Eqs. (12)—(14) would eliminate the restriction of observations to different and equally-spaced times. In addition, an appropriate combination of the techniques used for nonstationary, uncorrelated noise and for stationary, correlated noise may be employed in the event of nonstationary, correlated noise. The calculations would then become more complicated and, perhaps in some cases, prohibitive.

Inversion of $\hat{\Sigma}^{(c)}$ in 3 must now be accomplished. If the noise is nonstationary and uncorrelated and $\sigma^2(t) = \sum_{k=1}^{p} \nu_k T_k(t)$, $(\hat{\Sigma}^{(c)})^{-1}$ may be written in closed form as

$$(\hat{\Sigma}^{(c)})^{-1} = \operatorname{diag}\left(\frac{1}{\sum_{k=1}^{p} \hat{\nu}_k^{(c)} T_{k1}}, \frac{1}{\sum_{k=1}^{p} \hat{\nu}_k^{(c)} T_{k2}}, \ldots, \frac{1}{\sum_{k=1}^{p} \hat{\nu}_k^{(c)} T_{kn}} \right) \tag{15}$$

for $c = 1, 2, \ldots, c^*$.

The performance of this inversion has not yet been investigated for correlated noise.

A valid, yet simple, measure† of the change in $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ for 5 is $\max_{j=1,2,\ldots,p} \left| \frac{\hat{\beta}_j^{(c)} - \hat{\beta}_j^{(c-1)}}{\hat{\beta}_j^{(c)}} \right|$.

3. INCORPORATION OF A PRIORI INFORMATION

Let $\beta_1, \beta_2, \ldots, \beta_p$ vary randomly over all possible states of nature with means $\beta_1^*, \beta_2^*, \ldots, \beta_p^*$ and covariance matrix Γ of variances $\gamma_j^2 = \gamma_{jj}$ and covariances γ_{jk}. Introducing matrix notation yields

$$\beta^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \cdot \\ \cdot \\ \cdot \\ \beta_p^* \end{bmatrix}$$

and

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1p} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \cdot \\ \gamma_{1p} & \gamma_{2p} & & \gamma_{pp} \end{bmatrix}.$$

The a priori information, which exists from previous analysis, is knowledge of β^* and Γ.

---

†Depending upon the application, $\max_{i=1,2,\ldots,n} \left| \frac{\hat{Y}_i^{(c)} - \hat{Y}_i^{(c-1)}}{\hat{Y}_i^{(c)}} \right|$, $\sum_{j=1}^{p} \left| \frac{\hat{\beta}_j^{(c)} - \hat{\beta}_j^{(c-1)}}{\hat{\beta}_j^{(c)}} \right|$ or $\sum_{i=1}^{n} \left| \frac{\hat{Y}_i^{(c)} - \hat{Y}_i^{(c-1)}}{\hat{Y}_i^{(c)}} \right|$ may be a more desirable measure of this change.

Two quite reasonable methods to ensure that estimators of $\beta_1, \beta_2, \ldots, \beta_p$ will be consistent with the a priori information are: incorporation of $\underline{\beta}^*$ and Γ into the quadratic form to be minimized by the estimators; and employment of a linear combination $\Lambda\hat{\underline{\beta}}^{(c*)} + (I_p - \Lambda)\underline{\beta}^*$ (a special case of which is $\lambda\hat{\underline{\beta}}^{(c*)} + (1-\lambda)\underline{\beta}^*$, with $\Lambda = \lambda I_p$ and $0 < \lambda < 1$) as the vector of estimators, where Λ is a matrix which measures one's relative confidence in the data (as represented by $\hat{\underline{\beta}}^{(c*)}$) and the a priori information (as represented by $\underline{\beta}^*$). Suppose

$$Q^* = (\underline{Y} - X\underline{\beta})'(\hat{\Sigma}^{(c*)})^{-1}(\underline{Y} - X\underline{\beta}) + (\underline{\beta} - \underline{\beta}^*)'\Gamma^{-1}(\underline{\beta} - \underline{\beta}^*)$$

is the quadratic form that incorporates $\underline{\beta}^*$ and Γ and is to be minimized.† Then it may be observed that the corresponding estimators $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*$ which incorporate $\underline{\beta}^*$ and Γ are given by

$$\begin{aligned}
\hat{\underline{\beta}}^* &= (\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*)' \\
&= [X'(\hat{\Sigma}^{(c*)})^{-1}X + \Gamma^{-1}]^{-1}[X'(\hat{\Sigma}^{(c*)})^{-1}\underline{Y} + \Gamma^{-1}\underline{\beta}^*] \\
&= [X'(\hat{\Sigma}^{(c*)})^{-1}X + \Gamma^{-1}]^{-1}[X'(\hat{\Sigma}^{(c*)})^{-1}X\hat{\underline{\beta}}^{(c*)} + \Gamma^{-1}\underline{\beta}^*] \\
&= \hat{\Lambda}\hat{\underline{\beta}}^{(c*)} + (I_p - \hat{\Lambda})\underline{\beta}^*,
\end{aligned} \tag{16}$$

where $\hat{\Lambda}$ is the p-by-p matrix of elements $\hat{\lambda}_{jk}$ defined as

$$\hat{\Lambda} = [X'(\hat{\Sigma}^{(c*)})^{-1}X + \Gamma^{-1}]^{-1}[X'(\hat{\Sigma}^{(c*)-1}X]. \tag{17}$$

Note that $\hat{\underline{\beta}}^{(c*)}$ and $\hat{\Sigma}^{(c*)}$ may be replaced by $\hat{\underline{\beta}}^{(0)}$ and I_n, if it is desired to incorporate the a priori information into least squares rather than IWLS.

It might easily be shown that $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*$ would be the maximum likelihood estimators if Σ and $\hat{\Sigma}^{(c*)}$ were equal for both the noise and the coefficients being normally distributed.

---

†If $0 < \hat{\lambda} < 1$, then two other quadratic forms that incorporate $\underline{\beta}^*$ and Γ and might be minimized are:

$$Q^{**} = \hat{\lambda}(\underline{Y} - X\underline{\beta})'(\hat{\Sigma}^{(c*)})^{-1}(\underline{Y} - X\underline{\beta}) + (1-\hat{\lambda})(X\underline{\beta} - X\underline{\beta}^*)'(\hat{\Sigma}^{(c*)})^{-1}(X\underline{\beta} - X\underline{\beta}^*),$$

with the corresponding estimators $\hat{\beta}_1^{**}, \hat{\beta}_2^{**}, \ldots, \hat{\beta}_p^{**}$ that incorporate $\underline{\beta}^*$ and Γ being yielded by

$$\hat{\underline{\beta}}^{**} = (\hat{\beta}_1^{**}, \hat{\beta}_2^{**}, \ldots, \hat{\beta}_p^{**})' = \hat{\lambda}\hat{\underline{\beta}}^{(c*)} + (1-\hat{\lambda})\underline{\beta}^*; \text{ and}$$

$$Q^{***} = \hat{\lambda}(\underline{Y} - X\underline{\beta})'(\hat{\Sigma}^{(c*)})^{-1}(\underline{Y} - X\underline{\beta}) + (1-\hat{\lambda})(\underline{\beta} - \underline{\beta}^*)'\Gamma^{-1}(\underline{\beta} - \underline{\beta}^*),$$

with the corresponding estimators $\hat{\beta}_1^{***}, \hat{\beta}_2^{***}, \ldots, \hat{\beta}_p^{***}$ that incorporate $\underline{\beta}^*$ and Γ being yielded by

$$\underline{\beta}^{***} = (\hat{\beta}_1^{***}, \hat{\beta}_2^{***}, \ldots, \hat{\beta}_p^{***})' = [\hat{\lambda}X'(\hat{\Sigma}^{(c*)})^{-1}X + (1-\hat{\lambda})\Gamma^{-1}]^{-1}[\hat{\lambda}X'(\hat{\Sigma}^{(c*)})^{-1}\underline{Y} + (1-\hat{\lambda})\Gamma^{-1}\underline{\beta}^*]$$

One such $\hat{\lambda}$ is

$$\hat{\lambda} = \|\hat{\Lambda}\| / (\|\hat{\Lambda}\| + \|I_p - \hat{\Lambda}\|)$$

for $\|\hat{\Lambda}\|$ being a norm of $\hat{\Lambda}$ $\left(\text{e.g., } \left(\sum_{j=1}^{p}\sum_{k=1}^{p}\hat{\lambda}_{jk}^2\right)^{1/2}\right)$.

4. PROPER ARRANGEMENT OF COEFFICIENTS AND DATA

To denote the kth data channel, for $k=1, 2, \ldots, q$, prefix a k to the subscripts of the previously-introduced notation and obtain, in particular: $X_{k1}(t), X_{k2}(t), \ldots, X_{kp_k}(t), \beta_{k1}, \beta_{k2}, \ldots, \beta_{kp_k}, Y_k(t)$ and $\epsilon_k(t)$; $X_{kji}=X_{kj}(t_{ki})$ for $j=1, 2, \ldots, p_k$, $Y_{ki}=Y_k(t_{ki})$ and $\epsilon_{ki}=\epsilon_k(t_{ki})=Y_{ki}-\sum_{j=1}^{p_k}\beta_{kj}X_{kji}$ for $i=1, 2, \ldots, n_k$; and

$$X_{kj}=\begin{bmatrix} X_{kj1} \\ X_{kj2} \\ \vdots \\ X_{kjn_k} \end{bmatrix} \qquad \text{for } j=1, 2, \ldots, p_k,$$

$$\underset{\sim}{X}_k=(\underset{\sim}{X}_{k1}, \underset{\sim}{X}_{k2}, \ldots, X_{kp_k}),$$

$$\underset{\sim}{\beta}_k=\begin{bmatrix} \beta_{k1} \\ \beta_{k2} \\ \vdots \\ \beta_{kp_k} \end{bmatrix},$$

$$Y_k=\begin{bmatrix} Y_{k1} \\ Y_{k2} \\ \vdots \\ Y_{kn_k} \end{bmatrix},$$

$$\epsilon_k=\begin{bmatrix} \epsilon_{k1} \\ \epsilon_{k2} \\ \vdots \\ \epsilon_{kn_k} \end{bmatrix},$$

and

$$\Sigma_{kk}=\begin{bmatrix} \sigma_{k11} & \sigma_{k12} & \ldots & \sigma_{k1n_k} \\ \sigma_{k12} & \sigma_{k22} & \ldots & \sigma_{k2n_k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{k1n_k} & \sigma_{k2n_k} & \ldots & \sigma_{kn_kn_k} \end{bmatrix}.$$

The analogues of Eqs. (1) and (2) for the kth channel are now

$$Y_k(t) = \sum_{j=1}^{\nu_k} \beta_{kj} X_{kj}(t) + \epsilon_k(t) \text{ and} \tag{18}$$

$$\underline{Y}_k = \sum_{j=1}^{\nu_k} \beta_{kj} \underline{X}_{kj} + \underline{\epsilon}_k = X_k \underline{\beta}_k + \underline{\epsilon}_k \text{ for } k = 1, 2, \ldots, q. \tag{19}$$

Let

$$X = \begin{bmatrix} X_1 & 0 & 0 & \ldots & 0 \\ 0 & X_2 & 0 & \ldots & 0 \\ 0 & 0 & X_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & X_q \end{bmatrix},$$

$$\underline{\beta} = \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_q \end{bmatrix},$$

$$\underline{Y} = \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \vdots \\ \underline{Y}_q \end{bmatrix},$$

and

$$\underline{\epsilon} = \begin{bmatrix} \underline{\epsilon}_1 \\ \underline{\epsilon}_2 \\ \vdots \\ \underline{\epsilon}_q \end{bmatrix}$$

which has the effect of stacking the data and coefficients associated with channels 2, 3, . . ., q under the data and coefficients associated with channel 1. If σ_{hkij} is the covariance between ϵ_{hi} and ϵ_{kj} and the n_h-by-n_k matrix Σ_{hk} is

$$\Sigma_{hk}=\begin{bmatrix} \sigma_{hk11} & \sigma_{hk12} & \cdots & \sigma_{hk1n_k} \\ \sigma_{hk21} & \sigma_{hk22} & \cdots & \sigma_{hk2n_k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{hkn_h1} & \sigma_{hkn_h2} & \cdots & \sigma_{hkn_hn_k} \end{bmatrix}.$$

then the covariance matrix of $\epsilon_{11}, \epsilon_{12}, \ldots, \epsilon_{1n_1}, \epsilon_{21}, \epsilon_{22}, \ldots, \epsilon_{2n_2}, \ldots, \epsilon_{q1}, \epsilon_{q2}, \ldots, \epsilon_{qn_q}$ is

$$\Sigma=\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1q} \\ \Sigma'_{12} & \Sigma_{22} & \cdots & \Sigma_{2q} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \Sigma'_{1q} & \Sigma'_{2q} & \cdots & \Sigma_{qq} \end{bmatrix}.$$

Suppose that $\beta_{k_1j_1}, \beta_{k_2j_2}, \ldots, \beta_{k_rj_r}$ are the same (i.e., the same coefficient appears in the linear models for channels $k_1, k_2, \ldots, k_r$). All but one of them, for example $\beta_{k_1j_1}$, are superfluous and should be eliminated from consideration; and the pertinent columns of $X_{k_1}, X_{k_2}, \ldots, X_{k_r}$ and $\underline{Y}_{k_1}, \underline{Y}_{k_2}, \ldots, \underline{Y}_k$ (i.e., the pertinent data from channels $k_1, k_2, \ldots, k_r$) should be utilized in the estimation of $\beta_{k_1j_1}$. This may be accomplished in a simple (though not simple-appearing) manner, by: deleting $\beta_{k_2j_2}, \beta_{k_3j_3}, \ldots, \beta_{k_rj_r}$ from $\underline{\beta}$; replacing column numbers $\sum_{k=1}^{k_1-1} p_k+j_1$ in X by the sum of column numbers $\sum_{k=1}^{k_1-1} p_k+j_1$, $\sum_{k=1}^{k_2-1} p_k+j_2, \ldots, \sum_{k=1}^{k_r-1} p_k+j_r$ in X (which has the effect of alining the j_2th column of X_{k_2}, the j_3th column of X_{k_3}, . . ., the j_rth column of X_{k_r} under the j_1th column of X_{k_1} in the formation of the $\left(\sum_{k=1}^{k_1-1} p_k+j_1\right)$th column of X); and deleting column numbers $\sum_{k=1}^{k_2-1} p_k+j_2, \sum_{k=1}^{k_3-1} p_k+j_3, \ldots, \sum_{k=1}^{k_r-1} p_k+j_r$ from X. If more convenient, one of $\beta_{k_1j_1}, \beta_{k_2j_2}, \ldots, \beta_{k_rj_r}$ other than $\beta_{k_1j_1}$ may be used as the focal point for the alinement.

After all duplications among the β_{kj}'s have been treated as indicated, the coefficients and data from all channels are properly arranged; and a properly arranged analogue of Equation (2) is given, in the same form as before, by

$$\underline{Y}=X\underline{\beta}+\underline{\epsilon}. \tag{20}$$

A proper arrangement of the data from all channels has to be considered in determining the form of Σ and in estimating its necessary variances and covariances for IWLS. In particular, the requirement of nonstationary, uncorrelated noise for this proper arrangement means that not only ϵ_{ki} and ϵ_{kj}

(i.e., observations of the noise at any two times within any channel) must be uncorrelated, but also that ϵ_{hi} and ϵ_{kj} (i.e., observations of the noise at any two times in any two channels) must be uncorrelated.

One may view the $p < \sum_{k=1}^{q} p_k$ resulting columns of X as p vectors $\underline{X}_j$ of $n = \sum_{k=1}^{q} n_k$ observations X_{ji} on the p component signals $X_j(t)$, the p elements of $\underline{\beta}$ as the p coefficients β_j of $X_j(t)$, the n elements of $\underline{Y}$ as n observations Y_i on the composite signal $Y(t)$, the n elements of $\underline{\epsilon}$ as n observations ϵ_i on the noise $\epsilon(t)$, the n^2 elements of Σ as the n variances σ_i^2 of ϵ_i and n^2-n covariances σ_{hi} of ϵ_h and ϵ_i and Eq. (20) as the matrix alter ego of Eq. (1).

5. PARTITIONING, SELECTING AND APPORTIONING

For exposition purposes, it is desirable: to change the number of linearly independent component signals and corresponding unknown coefficients in the linear model to m; to redesignate them by $Z_1(t), Z_2(t), \ldots, Z_m(t)$ and $\alpha_1, \alpha_2, \ldots, \alpha_m$; and to transform $Z_k(t)$, via multiplication by a positive scale factor, so that the units of $Z_k(t)$ will become the units of $Y(t)$ and α_k will become unit-less for $k=1, 2, \ldots, m$. This transformation will be compensated for in the technique described.

Let $Z_{ki}=Z_k(t_i)$ for $i=1, 2, \ldots, n$ and $k=1, 2, \ldots, m$; $\alpha_1, \alpha_2, \ldots, \alpha_m$ vary randomly over the ensemble of all possible states of nature with means $\alpha_1^*, \alpha_2^*, \ldots, \alpha_m^*$ and covariance matrix Γ^* of variances $\gamma_k^{*2}=\gamma_{kk}^*$ and covariances γ_{hk}^*; and

$$\underline{Z}_k = \begin{bmatrix} Z_{k1} \\ Z_{k2} \\ \cdot \\ \cdot \\ \cdot \\ Z_{kn} \end{bmatrix} \qquad \text{for } k=1, 2, \ldots, m,$$

$$Z = (\underline{Z}_1, \underline{Z}_2, \ldots, \underline{Z}_m),$$

$$\underline{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_m \end{bmatrix}$$

$$\underline{\alpha}^* = \begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \cdot \\ \cdot \\ \cdot \\ \alpha_m^* \end{bmatrix},$$

and

$$\Gamma^* = \begin{bmatrix} \gamma_{11}^* & \gamma_{12}^* & \cdots & \gamma_{1m}^* \\ \gamma_{12}^* & \gamma_{22}^* & \cdots & \gamma_{2m}^* \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \gamma_{1m}^* & \gamma_{2m}^* & \cdots & \gamma_{mm}^* \end{bmatrix}$$

Then the linear model, its matrix alter ego and the corresponding a priori information become $\sum_{k=1}^{m} \alpha_k Z_k(t)$, $\sum_{k=1}^{m} \alpha_k \underline{Z}_k = Z\underline{\alpha}$ and $\underline{\alpha}^*$ and Γ^*.

When m is large and/or some of $Z_1(t), Z_2(t), \ldots, Z_m(t)$ are highly-related, consideration of accuracy and ease of computation frequently dictates that $\sum_{k=1}^{m} \alpha_k Z_k(t)$ be replaced by a new linear model containing only representative component signals which are highly-descriptive, but not highly-related. Suppose that the $p < m$ representative component signals, whose units are the units of $Y(t)$, and corresponding unit-less unknown coefficients are $X_1(t), X_2(t), \ldots, X_p(t)$ and $\beta_1, \beta_2, \ldots, \beta_p$; $X_{ji} = X_j(t_i)$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$; $\beta_1, \beta_2, \ldots, \beta_p$ vary randomly over the ensemble of all possible states of nature with means $\beta_1^*, \beta_2^*, \ldots, \beta_p^*$ and covariance matrix Γ of variances $\gamma_j^2 = \gamma_{jj}$ and covariances γ_{hj}; and

$$\underline{X}_j = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \cdot \\ \cdot \\ \cdot \\ X_{jn} \end{bmatrix} \qquad \text{for } j = 1, 2, \ldots, p,$$

$$X = (\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p)$$

$$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}$$

$$\underline{\beta}^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \cdot \\ \cdot \\ \cdot \\ \beta_p^* \end{bmatrix}$$

and

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1p} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \gamma_{1p} & \gamma_{2p} & \cdots & \gamma_{pp} \end{bmatrix}$$

Hence the new linear model, its matrix alter ego and the corresponding a priori information are $\sum_{j=1}^{p} \beta_j X_j(t)$, $\sum_{j=1}^{p} \beta_j \underline{X}_j = X\underline{\beta}$, and $\underline{\beta}^*$ and Γ.

One manner of replacing $\sum_{k=1}^{m} \alpha_k Z_k(t)$ by $\sum_{j=1}^{p} \beta_j X_j(t)$, and in particular $Z\underline{\alpha}$ by $X\underline{\beta}$, is to partition $Z_1(t), Z_2(t), \ldots, Z_m(t)$ into subsets $S_1, S_2, \ldots, S_p$ of highly-related $Z_k(t)$'s and to select an $X_j(t)$ to represent S_j (i.e., each $Z_k(t)$ in S_j) for $j = 1, 2, \ldots, p$. In addition, it is desirable to be able to apportion an estimator of each β_j among the α_k's which correspond to the $Z_k(t)$'s in S_j.

A measure of the degree of linear dependence between $Z_h(t)$ and $Z_k(t)$ is needed to accomplish this partitioning. One such measure, based upon $\underline{Z}_h$ and $\underline{Z}_k$, is the cosine

$$d_{hk} = d_{kh} = \frac{\sum_{i=1}^{n} Z_{hi} Z_{ki}}{\left[\left(\sum_{i=1}^{n} Z_{hi}^2\right)\left(\sum_{i=1}^{n} Z_{ki}^2\right)\right]^{1/2}} = \frac{\underline{Z}_h' \underline{Z}_k}{[(\underline{Z}_h' \underline{Z}_h)(\underline{Z}_k' \underline{Z}_k)]^{1/2}} \tag{21}$$

of the angle $D_{hk} = D_{kh}$ between $\underline{Z}_h$ and $\underline{Z}_k$. If

$$\bar{Z}_k = \frac{1}{n} \sum_{i=1}^{n} Z_{ki} \tag{22}$$

and

$$\bar{\underline{Z}}_k = \begin{bmatrix} \bar{Z}_k \\ \bar{Z}_k \\ \cdot \\ \cdot \\ \cdot \\ \bar{Z}_k \end{bmatrix} \quad \text{for } k = 1, 2, \ldots, m, \tag{23}$$

then the sample correlation coefficient

$$\begin{aligned} r_{hk} = r_{kh} &= \frac{\sum_{i=1}^{n} (Z_{hi} - \bar{Z}_h)(Z_{ki} - \bar{Z}_k)}{\left\{\left[\sum_{i=1}^{n} (Z_{hi} - \bar{Z}_h)^2\right]\left[\sum_{i=1}^{n} (Z_{ki} - \bar{Z}_k)^2\right]\right\}^{1/2}} \\ &= \frac{(\underline{Z}_h - \bar{\underline{Z}}_h)'(\underline{Z}_k - \bar{\underline{Z}}_k)}{\{[(\underline{Z}_h - \bar{\underline{Z}}_h)'(\underline{Z}_h - \bar{\underline{Z}}_h)][(\underline{Z}_k - \bar{\underline{Z}}_k)'(\underline{Z}_k - \bar{\underline{Z}}_k)]\}^{1/2}} \end{aligned} \tag{24}$$

between $\underline{Z}_h$ and $\underline{Z}_k$ is the cosine of the angle between $\underline{Z}_h - \bar{\underline{Z}}_h$ and $\underline{Z}_k - \bar{\underline{Z}}_k$ and also a measure of the degree of linear dependence between $Z_h(t)$ and $Z_k(t)$. Observe that the multiplication of $Z_h(t)$ by c_h and $Z_k(t)$ by c_k does not change the magnitude of d_{hk} and r_{hk} and changes the sign of d_{hk} and r_{hk} only for c_h and c_k having different signs. It might be demonstrated that any three of the d_{hk}'s and any three of the r_{hk}'s satisfy the inequalities

$$-1 < d_{hi}d_{ki} - [(1 - d_{hi}^2)(1 - d_{ki}^2)]^{1/2} \leqq d_{hk} \leqq d_{hi}d_{ki} + [(1 - d_{hi}^2)(1 - d_{ki}^2)]^{1/2} < 1 \tag{25}$$

and

$$-1 < r_{hi}r_{ki} - [(1 - r_{hi}^2)(1 - r_{ki}^2)]^{1/2} \leqq r_{hk} \leqq r_{hi}r_{ki} + [(1 - r_{hi}^2)(1 - r_{ki}^2)]^{1/2} < 1. \tag{26}$$

Consequently, it is feasible to utilize either d_{hk} or r_{hk} in a partitioning procedure for obtaining $S_1, S_2, \ldots, S_p$ from $Z_1(t), Z_2(t), \ldots, Z_m(t)$. The pertinent one to use is d_{hk} when a constant term is

not included in the linear model and r_{hk} when a constant term is included in the linear model. Without loss of generality, d_{hk} will be used in the text; but it should be replaced by r_{hk} where appropriate.

A reasonable partitioning procedure should not be affected by a renumbering of $Z_1(t), Z_2(t), \ldots, Z_m(t)$. The procedure also ought to yield $S_1, S_2, \ldots, S_p$ with a relatively high degree of linear dependence existing between any two $Z_k(t)$'s in the same S_j and a relatively low degree of linear dependence existing between any two $Z_k(t)$'s in different S_j's. Finally, it should provide some control over the size of p.

The partitioning procedure employed by EIWLS:

1. Computes

$$d_k = \sum_{h=1}^{k-1} |d_{hk}| + \sum_{h=k+1}^{m} |d_{hk}| = \sum_{h=1}^{m} |d_{hk}| - 1 \tag{27}$$

for $k = 1, 2, \ldots, m$.

2. Selects that $Z_k(t)$, say $Z_{k_1}(t)$, whose d_k is smallest.

3. Puts $Z_k(t)$ into S_1 if and only if $|d_{kk_1}| \geq d$, where $\cos 45° = 0.70711 \leq d < 1$ is a preassigned constant which should be selected to reflect the desire for the existence of a relatively high degree of linear dependence between any two $Z_k(t)$'s in the same S_j and a relatively low degree of linear dependence between any two $Z_k(t)$'s in different S_j's and the desire for a relatively small p.

4. Selects that $Z_k(t)$ not in S_1, say $Z_{k_2}(t)$, whose d_k is smallest among the $Z_k(t)$'s not in S_1.

5. For $Z_k(t)$ not in S_1, puts it into S_2 if and only if $|d_{kk_2}| \geq d$; and for $Z_k(t)$ in S_1, removes it from S_1 and puts it into S_2 if and only if $|d_{kk_2}| > |d_{kk_1}|$.

6. Continues in the same manner and selects that $Z_k(t)$ not in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1}, say $Z_{k_j}(t)$, whose d_k is smallest among the $Z_k(t)$'s not in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1}.

7. For $Z_k(t)$ not in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1}, puts it into S_j if and only if $|d_{kk_j}| \geq d$; and for $Z_k(t)$ in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1}, removes it from $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1} and puts it into S_j if and only if $|d_{kk_j}| > \max_{h=1,2,\ldots,j-1} |d_{kk_h}|$.

8. Continues in the same manner until selecting a $Z_k(t)$ not in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1}, say $Z_{k_p}(t)$, for which $|d_{kk_p}| \geq d$ for all the $Z_k(t)$'s not in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1} and, thereby, defining p.

9. For $Z_k(t)$ not in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1}, puts it into S_p; and for $Z_k(t)$ in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1}, removes it from $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1} and puts it into S_p if and only if

$$|d_{kk_p}| > \max_{j=1,2,\ldots,p-1} |d_{kk_j}|.$$

10. Inspects the resulting $S_1, S_2, \ldots, S_p$ to determine if $|d_{hk}|$ is sufficiently large for all $Z_h(t)$ and $Z_k(t)$ in the same S_j, if $|d_{hk}|$ is sufficiently small for all $Z_h(t)$ and $Z_k(t)$ in different S_j's and if p is sufficiently small.

11. Modifies $S_1, S_2, \ldots, S_p$ and d until 10 is satisfied to a reasonable extent.

It is notationally convenient to define $0 = m_0 < m_1 < m_2 < \ldots < m_{p-1} < m_p = m$ and renumber $Z_1(t), Z_2(t), \ldots, Z_m(t)$ with $Z_{m_{j-1}+1}(t), Z_{m_{j-1}+2}(t), \ldots, Z_{mj}(t)$ being in S_j for $j = 1, 2, \ldots, p$. Using Eq. (25), it might also be demonstrated that each $Z_k(t)$ in S_j may be transformed, via multiplication by $d_{m_j k}/|d_{m_j k}|$, so that d_{hk} becomes positive when $Z_h(t)$ and $Z_k(t)$ are in S_j for $j = 1, 2, \ldots, p$. This transformation will also be compensated for in the ensuing discussion.

The replacement of $\sum_{k=1}^{n} \alpha_k Z_k(t)$ by $\sum_{j=1}^{p} \beta_j X_j(t)$ means that $X_1(t), X_2(t), \ldots, X_p(t)$ should be selected to satisfy

$$\sum_{j=1}^{p} \beta_j X_j(t) = \sum_{k=1}^{m} \alpha_k Z_k(t). \tag{28}$$

In addition, the representation of S_j by $X_j(t)$ suggests that $X_j(t)$ should be selected to satisfy

$$\beta_j X_j(t) = \sum_{k=m_{j-1}+1}^{m_j} \alpha_k Z_k(t) \tag{29}$$

and

$$X_j(t) = \sum_{k=m_{j-1}+1}^{m_j} v_k Z_k(t), \tag{30}$$

where $v_{m_{j-1}+1}, v_{m_{j-1}+2}, \ldots, v_{m_j}$ are nonnegative weights whose sum is one, for $j=1, 2, \ldots, p$. Simple analytic consequences of the validity of Eqs. (29) and (30) for all values of t are

$$\alpha_k = v_k \beta_j \text{ for } k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j \tag{31}$$

and

$$\beta_j = \sum_{k=m_{j-1}+1}^{m_j} \alpha_k \text{ for } j=1, 2, \ldots, p. \tag{32}$$

Regardless of the true relationships (or lack of them) among $\alpha_{m_{j-1}+1}, \alpha_{m_{j-1}+2}, \ldots, \alpha_{m_j}$, the selecting of $X_j(t)$ to represent S_j (i.e., Eqs. (29) and (30)) induces analytic relationships among $\alpha_{m_{j-1}+1}, \alpha_{m_{j-1}+2}, \ldots, \alpha_{m_j}$ and β_j (i.e., Eqs. (31) and (32)) which cause them to become analytically and probabilistically indistinguishable (i.e., yield all of them once any one of them is determined) for $j=1, 2, \ldots, p$. The selecting, in turn, induces a web of analytic relationships, which are not difficult to derive, into the structure of the a priori information:

$$\alpha_k^* = v_k \beta_j^* \text{ for } k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j; \tag{33}$$

$$\beta_j^* = \sum_{k=m_{j-1}+1}^{m_j} \alpha_k^*; \tag{34}$$

$$\gamma_k^{*2} = \gamma_{kk}^* = v_k^2 \gamma_j^2 \text{ for } k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j; \tag{35}$$

$$\gamma_{hk}^* = v_h v_k \gamma_j^2 = \gamma_h^* \gamma_k^* \text{ for } k = h+1, h+2, \ldots, m_j \text{ and } h = m_{j-1}+1, m_{j-1}+2, \ldots, m_j - 1; \tag{36}$$

and

$$\gamma_j^2 = \left(\sum_{k=m_{j-1}+1}^{m_j} \gamma_k^* \right)^2 \text{ for } j=1, 2, \ldots, p. \tag{37}$$

Consideration of this structure produces

$$v_k = \gamma_k^* / \sum_{h=m_{j-1}+1}^{m_j} \gamma_h^* = \gamma_k^* / \gamma_j \text{ for } k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j \text{ and } j=1, 2, \ldots, p \tag{38}$$

as the natural set † of weights to use in Eqs. (30) and (31). It follows immediately that

$$\frac{\alpha_k - \alpha_k^*}{\gamma_k^*} = \frac{\beta_j - \beta_j^*}{\gamma_j};$$

and so α_k and β_j occur with equal probability when α_k, hence β_j, is normally distributed for $k = m_{j-1}+1$, $m_{j-1}+2$, . . ., m_j and $j = 1, 2, \ldots, p$.

Hence, the selecting of $X_1(t), X_2(t), \ldots, X_p(t)$ from $Z_1(t), Z_2(t), \ldots, Z_m(t)$ is obtained by

$$X_j(t) = \sum_{k=m_{j-1}+1}^{m_j} \left(\frac{\gamma_k^*}{\sum_{h=m_{j-1}+1}^{m_j} \gamma_h^*} \right) Z_k(t) \text{ for } j = 1, 2, \ldots, p; \tag{39}$$

the selecting of $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p$ from $\underline{Z}_1, \underline{Z}_2, \ldots, \underline{Z}_m$ is accomplished by

$$X_{ji} = \sum_{k=m_{j-1}+1}^{m_j} \left(\frac{\gamma_k^*}{\sum_{h=m_{j-1}+1}^{m_j} \gamma_h^*} \right) Z_{ki} \text{ for } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, p, \tag{40}$$

in particular; and the apportioning of an estimator $\hat{\beta}_j$ (e.g., $\hat{\beta}_j^{(0)}$, $\hat{\beta}_j^{(c*)}$ or $\hat{\beta}^*_j$) of β_j among $\alpha_{m_{j-1}+1}$, $\alpha_{m_{j-1}+2}$, . . ., α_{m_j} is provided by

$$\hat{\alpha}_k = \left(\frac{\gamma_k^*}{\sum_{h=m_{j-1}+1}^{m_j} \gamma_h^*} \right) \hat{\beta}_j \text{ for } k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j \text{ and } j = 1, 2, \ldots, p. \tag{41}$$

The combination of the two previous transformations to each $Z_k(t)$ in S_j (i.e., its multiplication by the product of the original $d_{m_jk}/|d_{m_jk}|$ and the corresponding positive scale factor) may now be compensated for by performing the same combination of transformations to the corresponding $\hat{\alpha}_k$ (i.e., its multiplication by the product of the original $d_{m_jk}/|d_{m_jk}|$ and the corresponding positive scale factor). Indeed, it might be proved that these transformations may be eliminated by replacing v_k in Eqs. (30), (31), (33), (35), (36), (39), (40), and (41) with the product of the original $d_{m_jk}/|d_{m_jk}|$, the corresponding positive scale factor and v_k.

6. DISCUSSION

A procedure has been described for estimation of the coefficients for linearly independent component signals in a linear model from observations on the component signals and a composite signal containing the linear model plus noise which is nonstationary and/or correlated with unknown covariance matrix, when the coefficients may vary randomly over all possible states of nature and they may appear in the linear models for more than one data channel, and when there may be a large number of component signals in the linear model and some of them may be highly-related. The procedure, denoted by EIWLS: properly arranges the coefficients and data from all channels; partitions

† An alternative set, which assigns equal weight to $Z_{m_{j-1}+1}, Z_{m_{j-1}+2}, \ldots, Z_{m_j}$ in Eq. (30) and to $\alpha_{m_{j-1}+1}, \alpha_{m_{j-1}+2}, \ldots, \alpha_{m_j}$ in Eq. (31), is given by $v_k = \frac{1}{m_j - m_{j-1}}$ for $k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j$ and $j = 1, 2, \ldots, p$.

the resulting set of component signals into subsets of highly-related ones and selects a representative component signal from each subset to include in the new linear model; obtains the least squares estimators of the coefficients in the new linear model, and then iteratively calculates an estimator of the noise covariance matrix and obtains the weighted least squares estimators of these coefficients which are determined by the inverse of the matrix estimator; incorporates the a priori information into estimators of the coefficients in the new linear model; and apportions an estimator for the coefficient of the representative component signal for a subset among the coefficients of component signals in that subset.

During the development of EIWLS, the difficulty of the generalized statistical regression problem and the requirement for an operational analysis within a reasonable period of time combined to dictate that the security of rigorous theoretical proof be occasionally sacrificed for the expedience of apparent theoretical implication plus intuitive justification. The computer provided a feasible means to implement the procedure and to empirically test it on examples when the noise is nonstationary and uncorrelated. Empirical testing of the procedure and its parts has included not only the analysis of fabricated examples (e.g., those in section 8 and Ref. [9]), but also the successful evaluation of inertial navigation systems via the analysis of real field-test data. Therefore, EIWLS is presented as a theoretically promising, intuitively appealing, and empirically tested (to a limited extent) solution for the generalized statistical regression problem.

The feasibility of IWLS depends upon the practical inversion of the estimators $\hat{\Sigma}^{(c)}$ of the noise covariance matrix Σ. This inversion has to be investigated in the case of correlated noise. When the noise is stationary and correlated and the observations are taken at different and equally-spaced times, the form of $\hat{\Sigma}^{(c)}$ is particularly simple and inversion may not be too difficult. A simplifying assumption, which is tenable in a large number of scientific problems, is that the noise covariance $\sigma(t, t')$ becomes zero as the time difference $|t-t'|$ increases beyond a certain limit. Then $\hat{\Sigma}^{(c)}$ has blocks of zero elements and inversion by partitioning may be practical.

Consider the IWLS estimators $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ of the coefficients $\beta_1, \beta_2, \ldots, \beta_p$ for $c=1, 2, \ldots, c^*$. In order to provide estimators of $\beta_1, \beta_2, \ldots, \beta_p$ of sufficient stability for a particular application, $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ need not converge to $\hat{\beta}_1^{(c^*)}, \hat{\beta}_2^{(c^*)}, \ldots, \hat{\beta}_p^{(c^*)}$, but only possess a valid measure of change (e.g., $\max_{j=1,2,\ldots,p} |(\hat{\beta}_j^{(c)}-\hat{\beta}_j^{(c-1)})/\hat{\beta}_j^{(c)}|$) which becomes less than a particular preassigned constant for $c=c_0, c_0+1, \ldots, c^*$. Suppose that $|(\hat{\beta}_j^{(c)}-\hat{\beta}_j^{(c-1)})/\hat{\beta}_j^{(c)}| \leq b_j$ for $c=c_j, c_j+1, \ldots, c^*$. Then $\hat{\beta}_j^{(c_j)}, \hat{\beta}_j^{(c_j+1)}, \ldots, \hat{\beta}_j^{(c^*)}$ appear to be approaching or oscillating in a band (whose width depends upon b_j) about an asymptotic value and could probably be combined, by some technique (e.g., extrapolation), to yield an improved estimator of β_j. This improved estimation of β_j should be investigated. A measure of the inherent accuracy in the estimation of β_j, which depends upon the matrix X of component signal observations and the vector $\underline{Y}$ of composite signal observations (and, implicitly, the vector $\underline{\epsilon}$ of noise observations) is provided by b_j and c_j. Note that there may be a considerable variation among the inherent accuracy in the estimation of $\beta_1, \beta_2, \ldots, \beta_p$.

At each iteration, the guiding principle of IWLS is to provide the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ and then Σ, based upon the available information at that iteration concerning Σ and then $\beta_1, \beta_2, \ldots, \beta_p$. This principle is similar to Bellman's principle of optimality in dynamic programming (Ref. [3], page 83), which states that "an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision." In addition, IWLS extracts considerably more information

from the data than does least squares and may be termed an "estimation servomechanism" or an "adaptive estimation procedure."

It is not at all unreasonable to conjecture that IWLS will produce, except perhaps under pathological conditions, near-optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ and Σ. However, the stability and estimation properties of IWLS should be further investigated, both theoretically and empirically.

It is informative to elaborate upon two previously mentioned characteristics of the optimum estimators for $\beta_1, \beta_2, \ldots, \beta_p$. The first characteristic is that they minimize the appropriate statistical distance $(Q_{\Sigma^{-1}})^{1/2}$ between the data vector $\underline{Y}$ and the estimator of the matrix alter ego for the linear model $X\underline{\beta}$. One has no knowledge whatsoever of the difference between the linear model

$$\sum_{j=1}^{p} \beta_j X_{ji}$$

at time t_i and its resulting estimator or of the difference between β_j and its estimator, but only knowledge of the probabilistic properties of these differences. The second characteristic is that for the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$, each of these differences has mean zero and minimum variance among all corresponding differences resulting from unbiased linear estimators of $\beta_1, \beta_2, \ldots, \beta_p$. In other words, the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ do not always produce an estimator which is closest to the quantity being estimated; but they do produce an estimator whose probability distribution is more tightly spread about this quantity than the probability distribution of any estimator resulting from unbiased linear estimators (e.g., the least squares estimators) of $\beta_1, \beta_2, \ldots, \beta_p$.

The three sets of estimators of $\beta_1, \beta_2, \ldots, \beta_p$ that incorporate the a priori information, one set given by Eq. (16) and the other two sets listed in the associated footnote, have been empirically tested on an example for which $\beta_1, \beta_2, \ldots, \beta_p$ had mean zero and covariance matrix $\Gamma = \text{diag}\,(\gamma_1^2, \gamma_2^2, \ldots, \ldots, \gamma_p^2)$ of variances $\gamma_1^2, \gamma_2^2, \ldots, \gamma_p^2$ that varied. Although this testing provided empirical justification for the use of Eq. (16), the results were somewhat inconclusive; since the version of the scalar weight presented in that footnote was too insensitive to the variation in $\gamma_1^2, \gamma_2^2, \ldots, \gamma_p^2$ and required an extreme variation in them to itself vary from zero to one. Additional empirical testing of the three sets of estimators that incorporate the a priori information, utilizing a more satisfactory version of $\hat{\lambda}$, is needed. The improved estimators of $\beta_1, \beta_2, \ldots, \beta_p$ and the appropriately modified version of $\hat{\Sigma}^{(c*)}$, rather than $\hat{\beta}_1^{(c*)}, \hat{\beta}_2^{(c*)}, \ldots, \hat{\beta}_p^{(c*)}$ and $\hat{\Sigma}^{(c*)}$, should be used to calculate these estimators.

Since the coefficients $\alpha_{mj-1+2}, \alpha_{mj-1+2}, \ldots, \alpha_{mj}$ of the component signals in the j^{th} subset S_j become analytically and probabilistically indistinguishable from the coefficient β_j of the representative component signal X_j for S_j for $j = 1, 2, \ldots, p$ during selecting and apportioning, the procedure used to partition the component signals $Z_1(t), Z_2(t), \ldots, Z_m(t)$ into subsets $S_1, S_2, \ldots, Sp$ is very important. Hence, the possibility of improving upon the partitioning procedure employed by EIWLS ought to be investigated. One approach would be to define a measure of the quality of a partition and to choose that procedure which maximized this measure.

7. SURVEY OF RELATED LITERATURE

In order to place EIWLS in proper perspective, a survey of related literature is now presented. This literature is grouped, for convenience, into seven categories of articles: those treating the properties of the least squares estimators, a set of weighted least squares estimators or the optimum esti-

mators (Refs. [12, 15, 22, 23, 26, 40–42]); those treating estimation of residuals or iterative estimation procedures (Refs. [6, 10, 21, 24, 29, 33–35, 37, 38]); those treating estimation procedures for correlated noise (Refs. [1, 14, 27, 28]); those treating estimation procedures which incorporate a priori information (Refs. [4, 7, 16, 19, 30, 36]); and those treating procedures for replacing the linear model by a new linear model and estimating the coefficients in the new linear model (Refs. [4, 7, 8, 10, 11, 20, 25, 39]).

Grenander and Rosenblatt (Ref. [15], sections 7.1–7.4) derive important asymptotic properties of the least squares estimators and the optimum estimators of the coefficients $\beta_1, \beta_2, \ldots, \beta_p$ when the the noise is stationary. The relationships among the columns of the component signal matrix X and the eigenvectors of the noise covariance matrix Σ and the conditions on the eigenvalues of Σ, required for the least squares estimators and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ to be equal, are obtained by Muller and Watson (Ref. [26]). Using the eigenvalues of certain matrices, Magness and McGuire (Ref. [23]) compare the covariance matrices of the least squares estimators and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$; and Golub (Ref. [12]) compares the covariance matrices of a set of weighted least squares estimators and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$.

Zyskind (Ref. [42]) and Watson (Ref. [40]) discuss the estimation problem in general; and necessary and sufficient conditions for the least squares and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ to be equal, or effectively equal, in particular. The contribution of errors in estimating weights to the variances of $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, for nonstationary and uncorrelated noise and for a special case of correlated noise, is treated by Williams (Ref. [41]). Both Williams (Ref. [41]) and McElroy (Ref. [22]) state necessary and sufficient conditions for the least squares and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ to be equal. The above asymptotic properties, relationships, conditions and comparisons should be useful in the investigation of procedures such as IWLS.

Material relevant to the estimation of regression residuals, and thereby Σ, is presented by Thiel (Refs. [35, 37]) and Koerts (Ref. [21]). When the noise is nonstationary and uncorrelated with variance

$$\sigma^2(t) = \left[\sum_{j=1}^{p} \beta_j X_j(t) \right]^r \sigma^2,$$

Prais and Aitchison (Ref. [29]) propose an iterative estimation procedure that is IWLS with the appropriate estimator $\hat{\Sigma}^{(c)}$ of the noise covariance matrix Σ; and Fisher (Ref. [10]) proposes two iterative solutions (modified forms of Newton's method of approximation) to the maximum likelihood equations resulting from normally distributed noise. Also analogous to IWLS is an iterative estimation procedure described by Turner, Monroe, and Lucas (Ref. [38]) for the case of the linear model being replaced by the quotient of two polynomials in the component signal $X(t)$ and the noise being stationary and uncorrelated. Mandel (Ref. [24]) develops an iterative estimation procedure, closely related to IWLS, to treat the linear model $\beta_1 + \beta_2 X(t)$ with nonstationary, uncorrelated noise. The convergence properties of $\hat{\beta}$ are treated by Telser (Ref. [34]).

Iterative estimation procedures which iterate over time, combining past data and estimators with new data, have appeared for some time in the engineering literature. Two early examples of such procedures are contained in Refs. [6, 33].

An excellent survey of articles concerned with estimation (and hypothesis testing) procedures for correlated noise is presented by Anderson (Ref. [1]). In the event of a linear model $\beta_1 + \beta_2 X(t)$ and stationary, correlated noise that is normally distributed with

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \cdots & \sigma^2\rho^{n-1} \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho^{n-2} \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \cdots & \sigma^2\rho^{n-3} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \sigma^2\rho^{n-1} & \sigma^2\rho^{n-2} & \sigma^2\rho^{n-3} & \cdots & \sigma^2 \end{bmatrix}$$

Murthy (Ref. [27]) discusses an iterative solution to the maximum likelihood equations and an explicit criterion for its convergence. Goodman (Ref. [14]) suggests a noniterative procedure, which performs the estimation in the frequency domain rather than in the time domain, when the noise is stationary and correlated. The estimation problem with a system of linear models for several data channels and noise correlated over both channels and time, is covered by Parks (Ref. [28]).

From two different points of view, Raiffa and Schlaifer (Ref. [30] sections 13.2–13.7) and Theil (Ref. [36]) cover the incorporation of a priori information into the estimation procedure for stationary, uncorrelated noise which is normally distributed with variance σ^2 either known or unknown. Gunckel (Ref. [16]) incorporates a priori information into estimators that reduce to the estimators $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*$ of $\beta_1, \beta_2, \ldots, \beta_p$ in Equation (16) when the noise has (mean zero and) covariance matrix $\hat{\Sigma}^{(c*)}$. The incorporation of a priori information described by Drucker (Ref. [7]) is in essentially the same form as $\underline{\hat{\beta}}^* = (\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*)'$. Chipman (Ref. [4]) presents material concerning the properties of $\underline{\hat{\beta}}^*$ (plus perhaps $\underline{\hat{\beta}}^{**}$ and $\underline{\hat{\beta}}^{***}$), as well as to partitioning, selecting and apportioning. Judge and Takayama (Ref. [19]) treat inequality restrictions, for incorporating a priori information into the estimation procedure.

Efroymson (Ref. [8]) proposes a stepwise selection of representative component signals via selection of: the best component signal $Z_k(t)$ to use as the representative component signal $X_1(t)$ in the linear model $\beta_1 X_1(t)$; the best $Z_k(t)$ to use with $X_1(t)$ as the representative component signal $X_2(t)$ in the linear model $\beta_1 X_1(t) + \beta_2 X_2(t)$; . . .; and the best $Z_k(t)$ to use with $X_1(t), X_2(t), \ldots, X_{p-1}(t)$ as the representative component signal $X_p(t)$ in the linear model

$$\sum_{j=1}^{p} \beta_j X_j(t).$$

A selection of representative component signals via an iterative nomination, of the best subset of $p - p'$ $Z_k(t)$'s to use with a subset $X_1(t), X_2(t), \ldots, X_{p'}(t)$ of p' $Z_k(t)$'s as the representative component signals $X_{p'+1}(t), X_{p'+2}(t), \ldots, X_p(t)$ in the linear model

$$\sum_{j=1}^{p} \beta_j X_j(t)$$

and then of the best subset of $p' Z_k(t)$'s to use with that subset $X_{p'+1}(t), X_{p'+2}(t), \ldots, X_p(t)$ of $p-p' Z_k(t)$'s as the representative component signals $X_1(t), X_2(t), \ldots, X_{p'}(t)$ in the linear model

$$\sum_{j=1}^{p} \beta_j X_j(t)$$

until two subsets renominate each other, is suggested by Villone, McCornack, and Wood (Ref. [39]). Both of these procedures simultaneously provide least squares estimators of $\beta_1, \beta_2, \ldots, \beta_p$. However, the two procedures which are most similar to partitioning, selecting, and apportioning are the procedure for approximating by representative component signals of Fisher (Ref. [10]) and the procedure for grouping, selecting, and apportioning of Drucker (Ref. [7]). Massy (Ref. [25]), Fortier and Solomon (Ref. [11]) and King (Ref. [20]) also are related to partitioning, selecting, and apportioning.

With the exception of Refs. [8, 39], this related literature became known to the author only after the development of EIWLS. One may, nevertheless, note the anticipation of: the iterative estimation procedures of Ref. [9] (IWLS) and the subsequent Ref. [24], by Refs. [10, 27, 29, 38]; the incorporation of a priori information into the estimation procedures of Ref. [9] (EIWLS in summarized form) and the subsequent Ref. [7], by Refs. [16, 30, 36]; and the procedure for partitioning, selecting and apportioning of Ref. [9] (EIWLS in summarized form) and the subsequent Ref. [7], by Ref. [10].

8. ILLUSTRATIVE EXAMPLE

An illustrative example concludes the paper. Although the example is described in detail by the paper's original version, only its input and output are summarized now for simplicity. For the example:

1. The 272 observations on each of the three channel linear models, $M_1(t)$, $M_2(t)$, and $M_3(t)$, are constructed from simultaneous observations on 14 known channel component signals using known channel coefficients.

2. The 272 observations on each of the three independent channel noises, $\epsilon_1(t)$, $\epsilon_2(t)$, and $\epsilon_3(t)$, are generated to be uncorrelated with mean zero and known variance $\sigma^2(t)$.

3. The 272 observations on each of the three channel composite signals, $Y_1(t)$, $Y_2(t)$, and $Y_3(t)$, are constructed by adding the appropriate channel linear model and noise observations.

4. The coefficients and data from all three channels are properly arranged, component signals are partitioned into seven subsets, and seven representative component signals are selected.

5. The coefficients, β_j, in the new linear model are estimated via least squares ($\hat{\beta}_j^{(0)}$), seven iterations of IWLS ($\hat{\beta}_j^{(7)}$), incorporation of a priori information ($\hat{\beta}_j^{*}$), and the optimum weighted least squares ($\hat{\beta}_{0j}$) for $j=1, 2, \ldots, 7$.

6. The corresponding estimators, $\hat{Y}_k^{(0)}(t)$, $\hat{Y}_k^{(7)}(t)$, $\hat{Y}_k^{*}(t)$ and $\hat{Y}_{0k}(t)$, of $Y_k(t)$, and $M_k(t)$ are computed for $k=1, 2, 3$.

Graphs of the three channel linear models and composite signals are presented in Figures 1, 2, and 3. Tables 1 and 2 show the actual coefficients in the channel and new linear models and their EIWLS estimators, and a summary of information regarding estimators of coefficients in the new linear model, respectively. The latter indicates that: (1) the seventh IWLS estimators are, mainly, both near-optimum and superior to the least squares estimators; (2) the majority of change in the IWLS estimators has occurred by the third iteration; and (3) the relative change between the sixth and seventh IWLS

estimators is quite small.* Actual comparisons of estimators of the three channel linear models are shown in Figures 4, 5, and 6. The superiority of the IWLS and EIWLS estimators over the least squares estimators is certainly corroborated by Figures 4, 5, and 6.

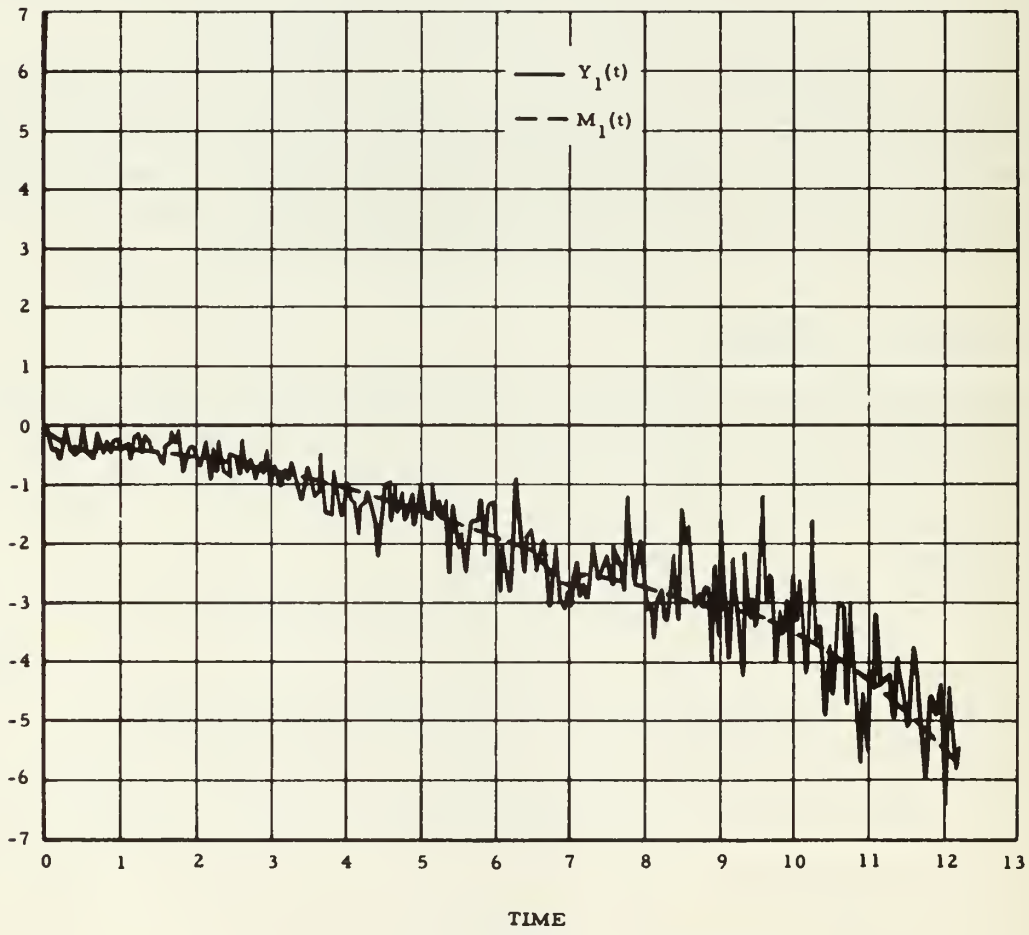


FIGURE 1. Composite signal and actual linear model for Channel 1.

*Dr. N. R. Mann recently observed that the $\hat{\beta}_j^{(1)}$'s are essentially closer to the $\hat{\beta}_j^{(0)}$'s than are the $\hat{\beta}_j^{(7)}$'s—which leads one to wonder to what the $\hat{\beta}_j^{(c)}$'s are converging, and whether IWLS and EIWLS should be stopped after the first iteration.

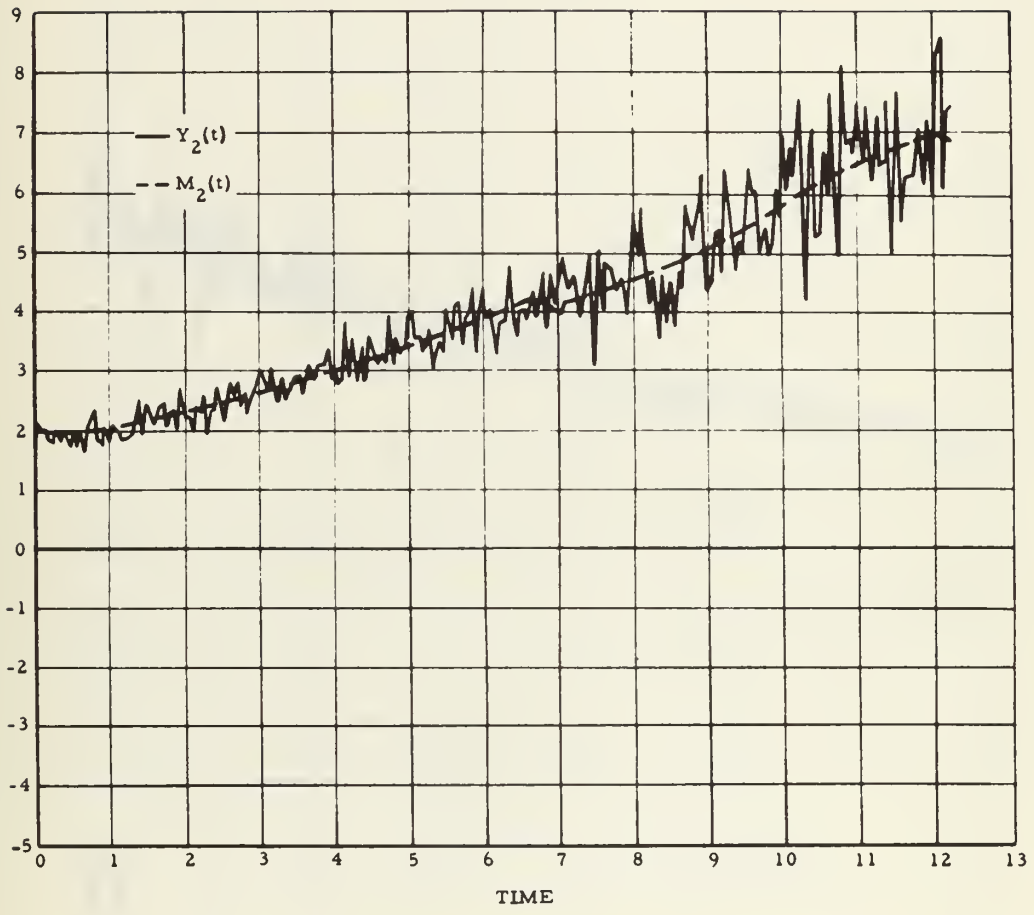


FIGURE 2. Composite signal and actual linear model for Channel 2.

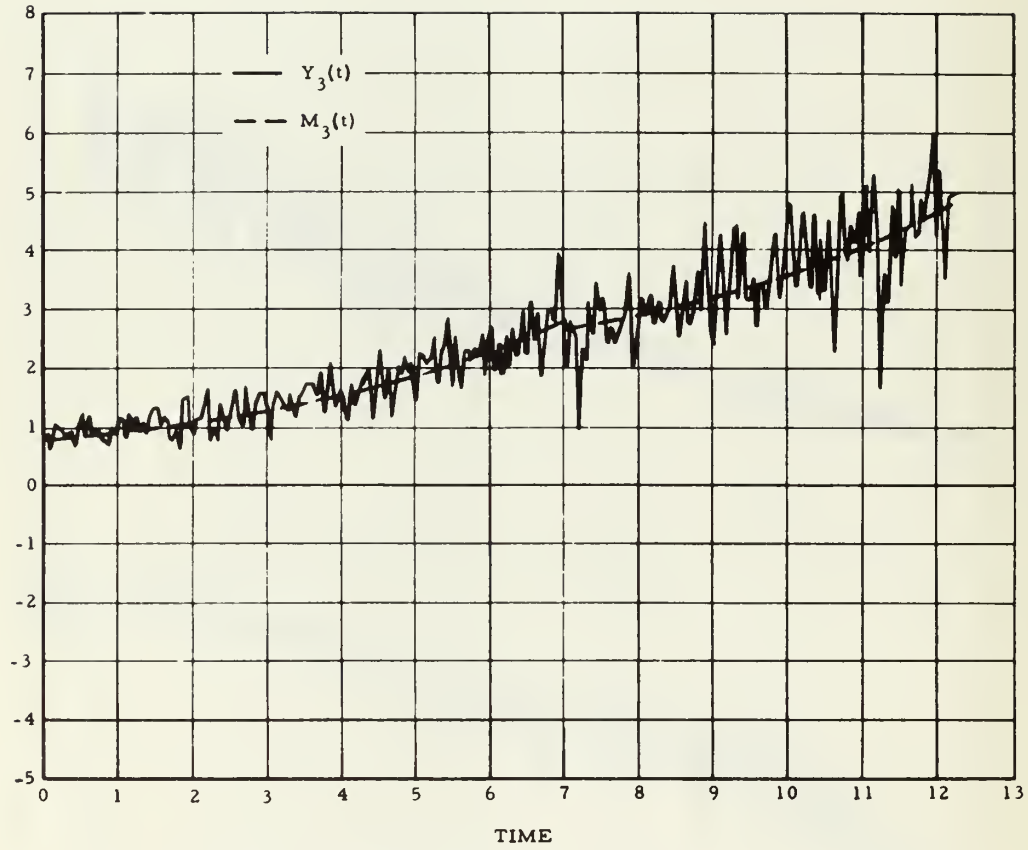


FIGURE 3. Composite signal and actual linear model for Channel 3.

TABLE 1. *Actual Coefficients in the Channel and New Linear Models and Their EIWLS Estimators*

Estimators	1	2	3	4	5	6	7	8	9	10	11	12	13	14
							Channel 1							
α_{1j}	−1.100	0.095	−1.600	0.450	0.030	−0.710	−0.150	−0.093	−0.029	0.060	0.027	0.014	0.018	−2.100
$\hat{\alpha}^*_{1j}$	−7.321	0.366	−0.732	−0.732	0.073	−0.732	−0.058	−0.070	−0.004	0.083	0.035	0.004	0.004	−3.518
							Channel 2							
α_{2j}	−2.200	0.020	−0.600	−0.700	−0.030	−0.920	−0.150	−0.093	−0.029	0.060	0.027	0.014	0.018	−2.100
$\hat{\alpha}^*_{2j}$	−6.698	0.340	−0.670	−0.670	−0.034	−0.670	−0.058	−0.070	−0.004	0.083	0.035	0.004	0.004	−3.518
							Channel 3							
α_{3j}	0.900	−0.085	0.300	0.750	−0.020	0.550	−0.150	−0.093	−0.029	0.060	0.027	0.014	0.018	−2.100
$\hat{\alpha}^*_{3j}$	5.846	−0.292	0.585	0.585	−0.029	0.585	−0.058	−0.070	−0.004	0.083	0.035	0.004	0.004	−3.518
							New [a]							
β_j	−2.835	−4.430	2.395	−0.150	−2.166	0.003	0.060							
$\hat{\beta}^*_j$	−9.957	−9.075	7.922	−0.058	−3.624	0.012	0.083							

[a] The association of α's with β's is as follows:

1. $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{16}$ with β_1
2. $\alpha_{21}, \alpha_{22}, \ldots, \alpha_{26}$ with β_2
3. $\alpha_{11}, \alpha_{32}, \ldots, \alpha_{36}$ with β_3
4. $\alpha_{17}, \alpha_{27}, \alpha_{37}$ with β_4
5. $\alpha_{18}, \alpha_{1.11}, \alpha_{1.14}, \alpha_{28}, \alpha_{2.11}, \alpha_{2.14}, \alpha_{38}, \alpha_{3.11}, \alpha_{3.14}$ with β_5
6. $\alpha_{19}, \alpha_{1.12}, \alpha_{1.13}, \alpha_{29}, \alpha_{2.12}, \alpha_{2.13}, \alpha_{39}, \alpha_{3.12}, \alpha_{3.13}$ with β_6
7. $\alpha_{1.10}, \alpha_{2.10}, \alpha_{3.10}$ with β_7

TABLE 2. *Summary of Information Regarding Estimators of Coefficients in the New Linear Model*

	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Optimum Estimators $\hat{\beta}_{0j}$	−9.5643	−9.1254	6.9091	−.078207	−3.5744	.028311	.085588
Least Squares Estimators $\hat{\beta}_j^{(0)}$	−9.7123	−7.5915	4.3828	−.169373	−3.5790	.086772	.079524
First IWLS Estimators $\hat{\beta}_j^{(1)}$	−9.9432	−8.7641	7.4216	−.076029	−3.6201	.024685	.082372
Second IWLS Estimators $\hat{\beta}_j^{(2)}$	−10.0185	−8.7293	7.6326	−.072770	−3.6311	.022528	.082753
Third IWLS Estimators $\hat{\beta}_j^{(3)}$	−9.9987	−8.7606	7.6502	−.071504	−3.6295	.021563	.082898
Fourth IWLS Estimators $\hat{\beta}_j^{(4)}$	−10.0012	−8.7444	7.6377	−.072353	−3.6301	.022033	.082857
Fifth IWLS Estimators $\hat{\beta}_j^{(5)}$	−9.9947	−8.7604	7.6455	−.071691	−3.6292	.021609	.082920
Sixth IWLS Estimators $\hat{\beta}_j^{(6)}$	−9.9976	−8.7499	7.6393	−.072178	−3.6298	.021893	.082879
Seventh IWLS Estimators $\hat{\beta}_j^{(7)}$	−9.9943	−8.7590	7.6438	−.071790	−3.6293	.021656	.082915
Estimators $\hat{\beta}_j^*$ that Incorporate A Priori Information	−9.9569	−9.0751	7.9218	−.058065	−3.6235	.011539	.082897
$\lvert(\hat{\beta}_j^{(0)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\rvert$	.0155	.1681	.3656	1.1657	.0013	2.0614	.0708
$\lvert(\hat{\beta}_j^{(7)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\rvert$	.0450	.0402	.1063	.0820	.0154	.2360	.0312
$\lvert(\hat{\beta}_j^{(7)} - \hat{\beta}_j^{(6)})/\hat{\beta}_j^{(7)}\rvert$	.0003	.0010	.0006	.0054	.0001	.0109	.0004

$\max_{j=1,2,\ldots,7} |(\hat{\beta}_j^{(0)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}| = 2.0614$ $\max_{j=1,2,\ldots,7} |(\hat{\beta}_j^{(7)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}| = .2360$ $\max_{j=1,2,\ldots,7} |(\hat{\beta}_j^{(7)} - \hat{\beta}_j^{(6)})/\hat{\beta}_j^{(7)}| = .0109$

$\sum_{j=1}^{7} \left|(\hat{\beta}_j^{(0)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right| = 3.8484$ $\sum_{j=1}^{7} \left|(\hat{\beta}_j^{(7)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right| = .5561$ $\sum_{j=1}^{7} \left|(\hat{\beta}_j^{(7)} - \hat{\beta}_j^{(6)})/\hat{\beta}_j^{(7)}\right| = .0187$

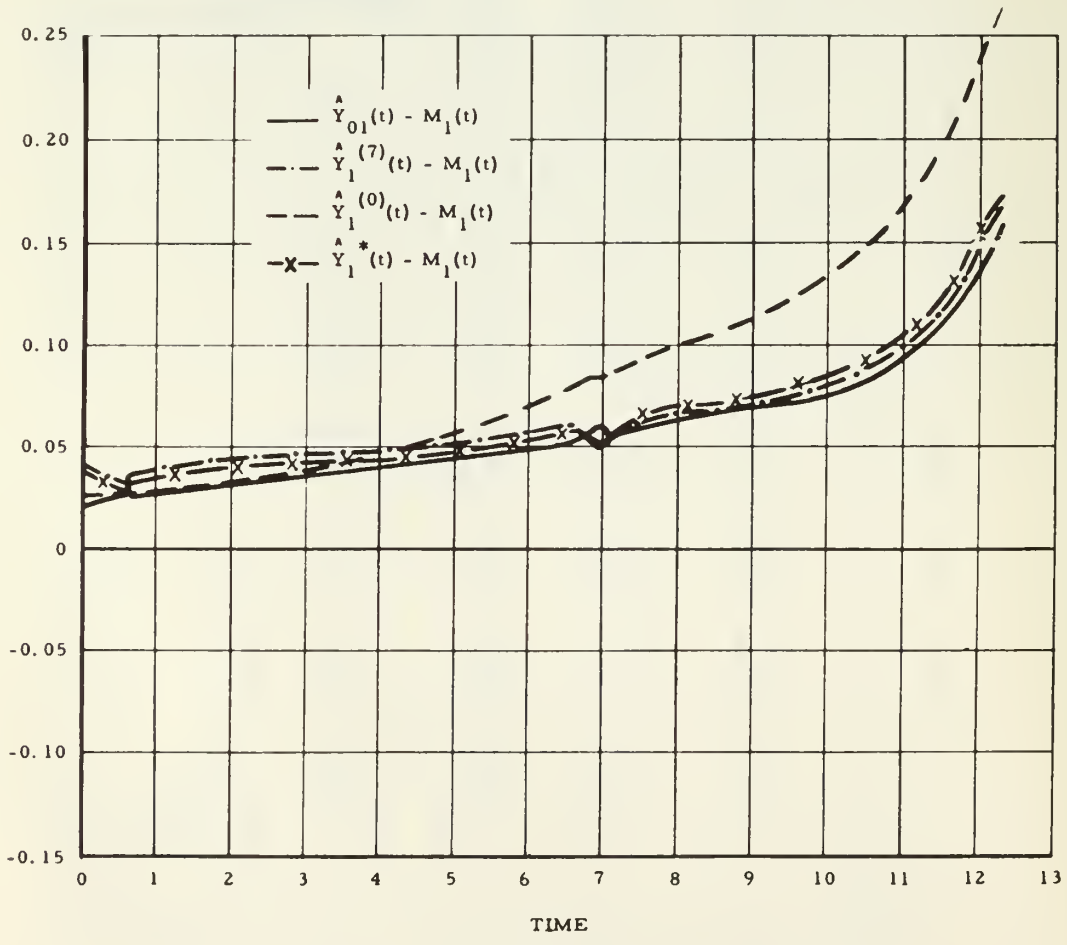


FIGURE 4. Actual comparison of estimators of the linear model for Channel 1.

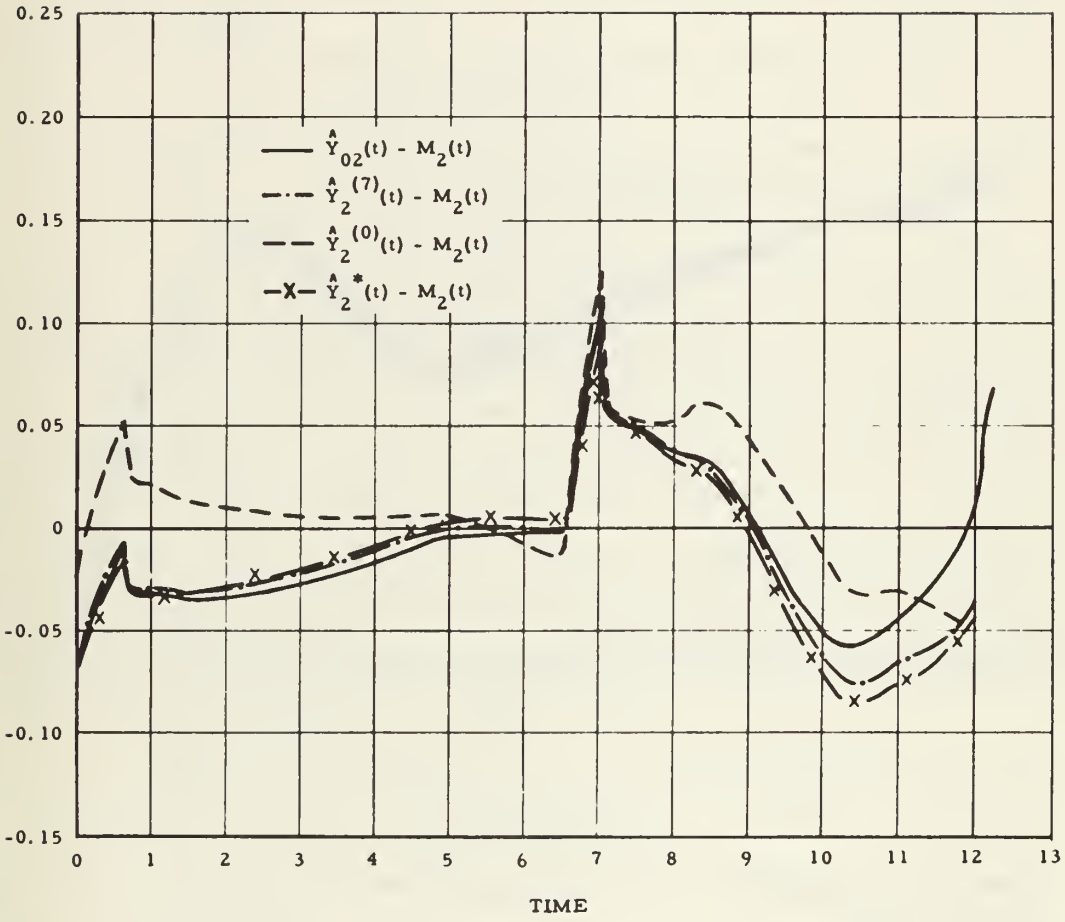


FIGURE 5. Actual comparison of estimators of the linear model for Channel 2.

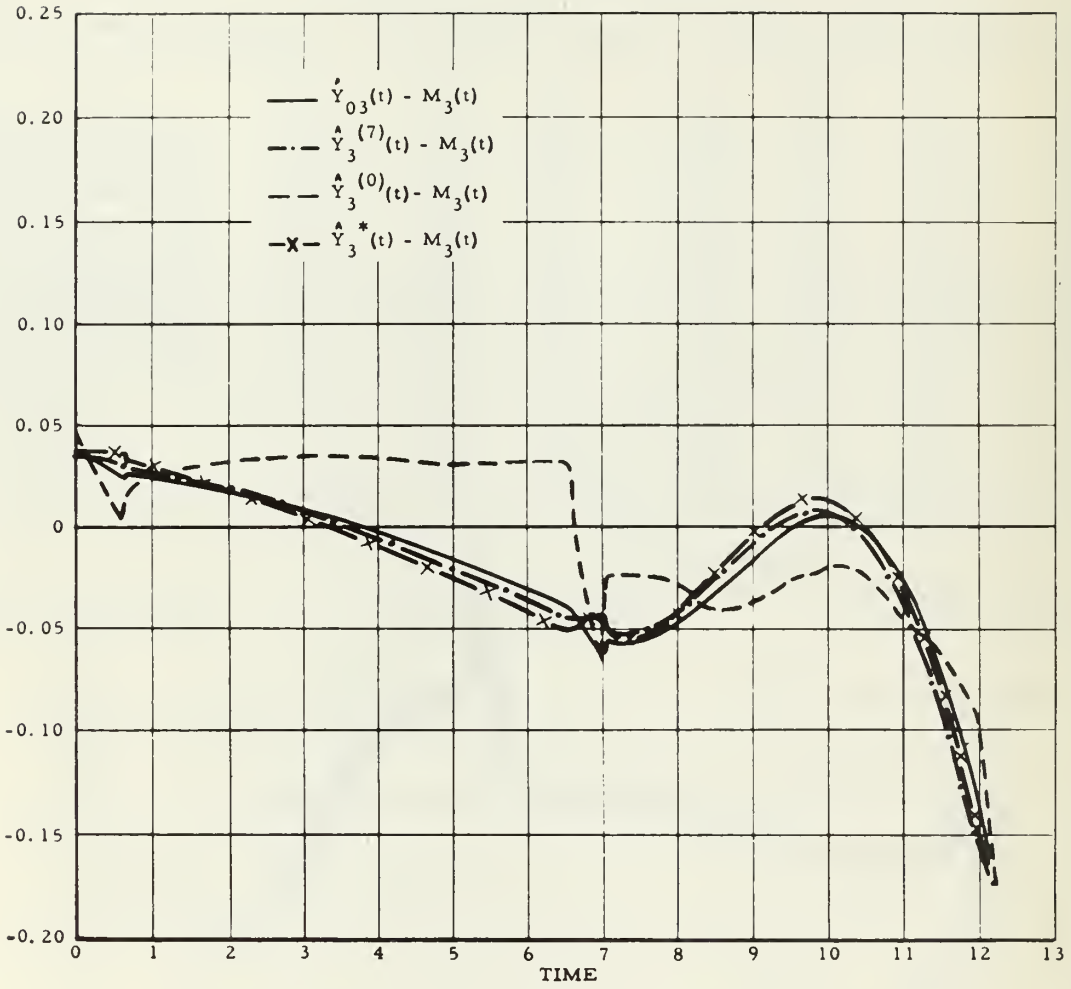


FIGURE 6. Actual comparison of estimators of the linear model for Channel 3.

REFERENCES

[1] Anderson, R. L., "The Problem of Autocorrelation in Regression Analysis," J. Am. Statis. Assoc. *49*, 113–129 (1954).

[2] Anderson, R. L., and T. A. Bancroft, *Statistical Theory in Research* (McGraw-Hill Book Company. New York, 1952).

[3] Bellman, R., *Dynamic Programming* (Princeton University Press, 1957).

[4] Chipman, J. S., "On Least Squares With Insufficient Observations," J. Am. Statis. Assoc. *59*, 1078–1111 (1964).

[5] Dalrymple, S. H., "NAWL4-North American Extended Iterative Weighted Least Squares 7094 FORTRAN Program," Autonetics Informal Document, North American Rockwell Corporation.

[6] Dennis, A. R., "An Adaptive Method for the Linear Estimation of Parameters in Discrete Non-stationary Time Series," TRW Systems Technical Memorandum 9990-6440-KU-000, TRW Inc. June (1963).

[7] Drucker, A. N., "The Determination of Tracking System Requirements and the Design of an Optimized Test Regime for Evaluating Ballistic Missile Inertial Guidance Systems in Flight Test,"

Advisory Committee to the Air Force Systems Command Tracking and Data Analysis Panel, National Academy of Sciences (Mar. 1964).

[8] Efroymson, M. A., "Multiple Regression Analysis," Chap. 17 of *Mathematical Methods for Digital Computers*, edited by A. Ralston and H. S. Wilf (John Wiley and Sons, Inc., 1960).

[9] Eisner, W., and A. F. Goodman, "Determination of Dominant Error Sources in an Inertial Navigation System by Iterative Weighted Least Squares," Autonetics Report T4–134/32, North American Rockwell Corporation (Mar. 1963), and American Institute of Aeronautics and Astronautics Journal *2*, 722–727 (1964).

[10] Fisher, G. R., "Iterative Solutions and Heteroscedasticity in Regression Analysis," Review of the International Statistical Institute *30*, 153–158 (1962).

[11] Fortier, J. J., and H. Solomon, "Clustering Procedures," *Proceedings of an International Symposium on Multivariate Analysis*, edited by P. R. Krishnaiah (Academic Press, 1966).

[12] Golub, G. H., "Comparison of the Variance of Minimum Variance and Weighted Least Squares Regression Coefficients," Ann. Math. Statist. *34*, 984–991 (1963).

[13] Goodman, A. F., "The Interface of Computer Science and Statistics," Western Division Paper 5204, McDonnell Douglas Astronautics Company (Feb. 1969).

[14] Goodman, N. R., Measuring Signals in the Presence of Stationary Noise with the Spectral Density of the Noise Unknown, Rocketdyne Research Report 62–20, North American Rockwell Corporation, October 1962.

[15] Grenander, U., and M. Rosenblatt, *Statistical Analysis of Stationary Time Series* (John Wiley and Sons, Inc., 1957).

[16] Gunckel, T. L., "The Estimation of Unknown Parameters Using Imperfect Information," Autonetics Inter-Office Letter 61–3243–4–PhA–58, North American Rockwell Corporation (Sept. 1961).

[17] Hsu, P. L., "NAWL2-North American Stepwise Iterative Weighted Least Squares 7090 or 7094 FORTRAN Program," Autonetics Technical Memorandum 243–2–157, North American Rockwell Corporation (Aug. 1963).

[18] Hsu, P. L., and K. V. Smith, "NAWL1-North American Iterative Weighted Least Squares 7090 or 7094 FORTRAN Program." Autonetics Technical Memorandum 243–2–131, North American Rockwell Corporation (Oct. 1962).

[19] Judge, G. G., and T. Takayama, "Inequality Restrictions in Regression Analysis," J. Am. Statist. Assoc. *61*, 166–181 (1966).

[20] King, B., "Step-Wise Clustering Procedures," J. Am. Statist. Assoc. *62*, 86–101 (1967).

[21] Koerts, J., "Some Further Notes on Disturbance Estimates in Regression Analysis," J. Am. Statist. Assoc. *62*, 169–183 (1967).

[22] McElroy, F. W., "A Necessary and Sufficient Condition that Ordinary Least-Squares Estimators Be Best Linear Unbiased," J. Am. Statist. Assoc. *62*, 1302–1304 (1967).

[23] Magness, T. A., and J. B. McGuire, "Comparison of Least Squares and Minimum Variance Estimates of Regression Parameters," Ann. Math. Statist. *33*, 462–470 (1962).

[24] Mandel, J., "Estimation of Weighting Factors in Linear Regression and Analysis of Variance," Technometrics *6*, 1–26 (1964).

[25] Massy, W. F., "Principal Components Regression in Exploratory Statistical Research," J. Am. Statist. Assoc. *60*, 234–256 (1965).

[26] Muller, M. E., and G. S. Watson, "Randomization and Linear Least Squares Estimation," Statistical Techniques Research Group Technical Report 32, Princeton University (July 1959).

[27] Murthy, V. K., "On Fitting a Linear Trend and Testing Independence When the Residuals Form a Markov Process," Applied Mathematics and Statistics Laboratories Technical Report 6, Grant DA-ORD-43, Stanford University (Nov. 1960).

[28] Parks, R. W., "Efficient Estimation of a System of Regression Equations When Disturbances are Both Serially and Contemporaneously Correlated," J. Am. Statist. Assoc. *62*, 500–509 (1967).

[29] Prais, S. J., and J. Aitchison, "The Grouping of Observations in Regression Analysis," Review of the International Statistical Institute *22*, 1–22 (1954).

[30] Raiffa, H., and R. Schlaifer, *Applied Statistical Decision Theory* (Harvard Business School, 1961).

[31] Rutemiller, H. C., and D. A. Bowers, "Estimation in a Heteroscedastic Regression Model," J. Am. Statist. Assoc. *63*, 552–557 (1968).

[32] Scheffe, H., *The Analysis of Variance* (John Wiley and Sons, Inc., 1959).

[33] "Stochastic Optimization of Ship's Inertial Navigation System," Autonetics Report C4-505/32, North American Rockwell Corporation (Apr. 1964).

[34] Telser, L. G., "Iterative Estimation of a Set of Linear Regression Equations," J. Am. Statist. Assoc. *59*, 845–863 (1964).

[35] Theil, H., "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances," J. Am. Statist. Assoc. *63*, 242–251 (1968).

[36] Theil, H., "On the Use of Incomplete Prior Information in Regression Analysis," J. Am. Statist. Assoc. *58*, 401–414 (1963).

[37] Theil, H., "The Analysis of Disturbances in Regression Analysis," J. Am. Statist. Assoc. *60*, 1067–1079 (1965).

[38] Turner, M. E., R. J. Monroe, and H. L. Lucas, "Generalized Asymptotic Regression and Non-Linear Path Analysis," Biometrics *17*, 120–143 (1961).

[39] Villone, L., R. L. McCornack, and K. R. Wood, "Multiple Regression with Subsetting of Variables," System Development Corporation Field Note FN-6622, System Development Corporation (June 1962); also "A Comparison of Three Predictor Selection Techniques in Multiple Regression, Psychometrika *35*, 257–271 (1970).

[40] Watson, G. S., "Linear Least Squares Regression," Ann. Math. Statist. *38*, 1679–1699 (1967).

[41] Williams, J. S., "The Variance of Weighted Regression Estimators," J. Am. Statist. Assoc. *62*, 1290–1301 (1967).

[42] Zyskind, G., "On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models," Ann. Math. Statist. *38*, 1092–1109 (1967).

THE LOWER BOUNDED AND PARTIAL UPPER BOUNDED DISTRIBUTION MODEL

A. Charnes, Fred Glover

and

D. Klingman

The University of Texas at Austin

1.0 Introduction

This paper considers a distribution model with a limited total supply and a lower bound on each origin's supply and each destination's demand. Our problem differs from the classical distribution problem in which supplies are, by contrast, bounded from above; however, like the classical transportation problem, this problem has a variety of practical interpretations, several of which are indicated below. This seemingly innocent difference gives rise to some paradoxical solution properties (see Ref. [2]), and has in fact occasioned an incorrect treatment of the problem in two highly reputable texts on linear programming and network flows (see Refs. [2, 3]).

We shall indicate an appropriate way to accommodate the lower bounded and partial upper bounded distribution model by proving that it is equivalent to a certain bounded variable distribution problem. The mechanism effecting this equivalence is related to, but slightly different from, that which transforms an ordinary distribution problem with unequal total supply and demand into another ordinary distribution problem in which total supply and demand are equal.

We describe the lower bounded and partial upper bounded distribution problem more precisely as follows. A total supply of N units is available for shipment. There are m origins, 0_i $i \epsilon I = \{1, 2, \ldots, m\}$, each of which must distribute at least a_i units of the total supply. There are n destinations, D_j $j \epsilon J = \{1, 2, \ldots, n\}$, each of which demands at least b_j units of the total supply. The objective is to minimize the total shipping cost where the cost of shipping from origin 0_i to destination D_j is denoted by c_{ij}. It is assumed that $N \geqq \Sigma\, a_i$, $N \geqq \Sigma\, b_j$, and that all supplies and demands are nonnegative.

1.1 EQUIVALENCE

In standard linear programming formulation, the lower bounded and partial upper bounded distribution model may be written as follows:

PROBLEM 1:

(1.1) $$\text{minimize} \sum_{\substack{i \epsilon I \\ j \epsilon J}} c_{ij} x_{ij}$$

subject to

(1.2) $$\sum_{j \epsilon J} x_{ij} \geqq a_i \qquad i \epsilon I,$$

(1.3) $$\sum_{i \epsilon I} x_{ij} \geqq b_j \qquad j \epsilon J,$$

$$\sum_{\substack{i\in I\\ j\in J}} x_{ij} \leqq N, \tag{1.4}$$

and

$$x_{ij} \geqq 0 \qquad i\in I,\ j\in J, \tag{1.5}$$

where $N \geqq \Sigma\ a_i$, $N \geqq \Sigma\ b_j$, a_i, and b_j are nonnegative, and x_{ij} denotes the amount shipped from origin i to destination j.

We now give a bounded variable problem that we shall prove to be equivalent to Problem 1.

PROBLEM 2:

$$\text{minimize} \sum_{\substack{i\in I'\\ j\in J'}} c_{ij}x_{ij} \tag{2.1}$$

subject to

$$\sum_{j\in J'} x_{ij} = a_i + N_I = a_i' \qquad i\in I \tag{2.2}$$

$$\sum_{i\in I'} x_{ij} = b_j + N_J = b_j' \qquad j\in J \tag{2.3}$$

$$\sum_{i\in I'} x_{i,\,n+1} = mN_I = b_{n+1}' \tag{2.4}$$

$$\sum_{j\in J'} x_{m+1,\,j} = \sum_{j\in J} b_j + nN_J - \sum_{i\in I} a_i = a_{m+1}' \tag{2.5}$$

$$0 \leqq x_{i,\,n+1} \leqq N_I \qquad i\in I' \tag{2.6}$$

$$0 \leqq x_{m+1,\,j} \leqq N_J \qquad j\in J \tag{2.7}$$

$$x_{ij} \geqq 0 \qquad i\in I,\ j\in J \tag{2.8}$$

where $I' = \{1, 2, \ldots, m+1\}$, $J' = \{1, 2, \ldots, n+1\}$, $N_I = \left(N - \sum_{i\in I} a_i\right)$, $N_J = \left(N - \sum_{j\in J} b_j\right)$, $c_{i,\,n+1} = 0$ $i\in I'$, $c_{m+1,\,j} = 0$ $j\in J$, and all the other parameters are the same as those of Problem 1.

THEOREM: The solution $x_{ij} = \bar{x}_{ij}$ $i\in I$, $j\in J$ is feasible (optimal) for Problem 1 if and only if the solution $x_{ij} = \bar{x}_{ij}$ $i\in I$, $j\in J$; $x_{i,\,n+1} = a_i' - \sum_{j\in J} \bar{x}_{ij}$ $i\in I$; $x_{m+1,\,j} = b_j' - \sum_{i\in I} \bar{x}_{ij}$ $j\in J$; and $x_{m+1,\,n+1} = \sum_{\substack{i\in I\\ j\in J}} \bar{x}_{ij} - \sum_{i\in I} a_i$ is feasible (optimal) for Problem 2. (Furthermore, the theorem is valid if the objective is to maximize the functionals (1.1) and (2.1).)

PROOF: It is apparent that the functionals for Problems 1 and 2 will have the same value for the solutions as indicated. Thus, to prove the theorem, it suffices simply to prove the feasibility assertion.

First, assume that the solution $x_{ij} = \bar{x}_{ij}$ $i\in I'$, $j\in J'$, is feasible for Problem 2. We will show that $x_{ij} = \bar{x}_{ij}$ $i\in I$, $j\in J$ is a feasible solution to Problem 1.

By assumption $\bar{x}_{i,\,n+1}$ satisfies (2.6). Moreover, from

$$\sum_{j\in J'} \bar{x}_{ij} = \sum_{j\in J} x_{ij} + \bar{x}_{i,\,n+1} = a_i + \left(N - \sum_{i\in I} a_i\right) i\in I,$$

we have $\sum_{j\in J} x_{ij} \geqq a_i$ $i\in I$. Consequently, (1.2) holds (for $x_{ij} = \bar{x}_{ij}$). The relation (1.3) is verified similarly.

Adding the Eqs. (2.2) and then subtracting (2.4) from the resulting equation, we obtain

$$\sum_{\substack{i\in J\\ j\in J}} \bar{x}_{ij} - \bar{x}_{m+1,\,n+1} = \sum_{\substack{i\in I\\ j\in J}} x_{ij} - \bar{x}_{m+1,\,n+1} = \sum_{i\in I} a_i.$$

where $\bar{x}_{m+1,\,n+1}$ satisfies (2.6) by hypothesis. Thus (1.4) is also satisfied. Also, (2.8) immediately implies (1.5). This completes the first half of the proof.

Next, assume $x_{ij}=\bar{x}_{ij}\ i\in I,\ j\in J$ is feasible for Problem 1. It is immediately apparent that (2.8) is satisfied by the solution specified in the theorem for Problem 2. Also, (2.2), (2.3), (2.4), and (2.5) follow directly from (1.2), (1.3), and the indicated values of $x_{i,\,n+1}$ and $x_{m+1,\,j}$. Since $x_{ij}=\bar{x}_{ij}\ i\in I,\ j\in J$ satisfies (1.2) and (1.4), the indicated values for $x_{i,n+1}\ i\in I$ and $x_{m+1,n+1}$ satisfy (2.6). (2.7) follows similarly from (1.3), completing the proof.

In reformulating Problem 1 into its equivalent form, Problem 2, we have introduced $(n+m+1)$ additional variables, which can be interpreted as adding a dummy origin and destination. By doing this, we are now able to solve Problem 1, for example, using the stepping-stone algorithm for a bounded variable distribution problem with $(m+1)$ origins and $(n+1)$ destinations.

1.2 APPLICATIONS

The model is susceptible to a variety of interesting practical applications. A few of these are as follows:

(a) Suppose that goods and services are optimally allocated in a certain economy according to a distribution model with supplies at each origin of a_i units and demands at each destination of b_j units (where $\Sigma b_j=\Sigma a_i$ and the shipping costs are all nonnegative). By utilizing the lower bounded and partial upper bounded distribution model, the economist is able to answer the question, "Given an unsaturated demand, if there is available $(N-\Sigma a_i)$ additional units of supply, how many of these additional units should be shipped, and along what origin-destination routes, to minimize total cost?"* In particular, solving Problem 2 for different values of N provides the economist with a curve indicating the relationship between total supply and the cost of shipping at least a_i units from each origin and receiving at least b_j units at each destination.

(b) If Problem 1 is changed by altering (1.4) to $\sum_{\substack{i\in I\\ j\in J}} x_{ij}=N$, then the only change required in Problem 2† is to replace (2.6) by

$$x_{m+1,\,n+1}=N_I$$

and

$$0\leqq x_{i,\,n+1}\leqq N_I\, i\in I.$$

---

*One might be tempted to suspect that shipping exactly $\Sigma a_i=\Sigma b_j$ units would yield the minimum cost, since the shipping costs are all nonnegative. This is not, however, the case. See [1] and [2].

†To verify this statement, observe that if $x_{m+1,\,n+1}=N_I$ then (2.4) implies $\sum_{i\in I} x_{i,\,n+1}=(m-1)N_I$. Adding the Eqs. of (2.2), we obtain $\sum_{\substack{j\in J'\\ i\in I}} x_{ij}=\Sigma\, a_i+mN_I$, and hence $\sum_{\substack{i\in I\\ j\in J}} x_{ij}=N$. Consequently, the new Problem 2 ships exactly N units through the set of cells $(i,j)\ i\in I,\ j\in J$.

Solving the new Problem 2 for different values of N (interpreting N as the total demand) provides the economist with a curve indicating the relationship between total demand and the cost of shipping at least a_i units from each origin and at least b_j units to each destination (see footnote on p. 279).

(c) This model could also be used by the government budget council in the allocation of funds to poverty organizations, such as welfare, Job Corps, Operation Headstart, etc., for the following scenario.

Suppose for certain political and economical reasons, it is necessary that the council allocate at least an amount of money, a_i, to each of the indicated organizations. Suppose these organizations, in turn, must allocate at least an amount of money, b_j, to the various socially and economically depressed areas in the United States to enable them to subsist at some designated minimal level.

Let c_{ij} represent the cost to government organization, i, for distributing, processing, and transferring one dollar to area j. Given a total of N dollars, ($N > \Sigma\, a_i$ and $N > \Sigma\, b_j$) appropriated for aiding these poverty areas, the budget council's question is "How much of the appropriation should be allocated in order to satisfy the minimum requirements of all groups concerned and at the same minimize the total cost incurred by the organization?"

1.3 OTHER DISTRIBUTION PROBLEMS

In this concluding section, we indicate two distribution problems which are special cases of Problem 1.

It may be observed that:

PROBLEM 3:

(3.1) $$\text{minimize} \sum_{\substack{i\epsilon I \\ j\epsilon J}} c_{ij}x_{ij},$$

subject to

(3.2) $$\sum_{j\epsilon J} x_{ij} \geqq a_i \qquad i\epsilon I, \qquad a_i \geqq 0,$$

(3.3) $$\sum_{i\epsilon I} x_{ij} = b_j \qquad j\epsilon J, \qquad b_j \geqq 0,$$

and

(3.4) $$x_{ij} \geqq 0 \qquad i\epsilon I,\ j\epsilon J, \qquad c_{ij} \geqq 0,$$

where $\Sigma\, a_i \leqq \Sigma\, b_j$ is a special case of Problem 1 with $N = \Sigma\, b_j$. Thus, while Problem 3 has previously been solved by modifying the stepping-stone method to handle negative slacks,* it can now be solved by the stepping-stone method for a bounded variable distribution problem using the equivalent form given by Problem 2.

---

*It may also be observed that the problem in which (3.2) is required to hold as an equality, (3.3) is permitted to hold as $\geqq$, and $\Sigma a_i \geqq \Sigma b_j$, is a special case of Problem 1. See [3] page 250 for an algorithm to solve these problems by a modified stepping-stone method.

We shall also show that:

PROBLEM 4:

$$\text{minimize} \sum_{\substack{i\epsilon I \\ j\epsilon J}} c_{ij}x_{ij},$$

subject to

$$\begin{aligned} \sum_{j\epsilon J} x_{ij} &\geqq a_i \qquad i\epsilon I \quad , a_i \geqq 0 \\ \sum_{i\epsilon I} x_{ij} &\geqq b_j \qquad j\epsilon J \quad , b_j \geqq 0 \\ x_{ij} &\geqq 0 \qquad i\epsilon I, j\epsilon J \quad , c_{ij} \geqq 0 \end{aligned}$$

is a special case of Problem 1.*

Since all $c_{ij} \geqq 0$, it is apparent that Problem 4 will have some finite optimal solution, say $x_{ij} = \bar{x}_{ij}$ $i\epsilon I$, $j\epsilon J$. Define $M = \sum_{\substack{i\epsilon I \\ j\epsilon J}} \bar{x}_{ij}$ (i.e., M is the total amount of goods shipped by the optimal solution). Then, letting N be any finite value $\geqq M$ in Problem 1, an optimal solution of Problem 1 is also optimal for Problem 4. In general, an acceptable value of N which will guarantee Problem 4 is a special case of Problem 1 is given by $N = m\left[\max_{i\epsilon I}\left(a_i, \sum_j b_j\right)\right]$. This follows directly from the fact that it is unnecessary to ship more than $\max(a_i, \Sigma b_j)$ units from origin 0_i in an optimal solution to Problem 4.

*For a discussion of the incorrect treatment in the literature of Problem 4, see [2].

REFERENCES

[1] Charnes, A., Fred Glover, and D. Klingman, "A Note on a Distribution Problem," Applied Mathematics for Management No. AMM–6, WP 69–19, University of Texas, Austin (1969).

[2] Charnes, A., and D. Klingman, "The 'More For Less' Paradox in the Distribution Model," Applied Mathematics for Management No. AMM–5, WP 69–18, University of Texas, Austin (1969).

[3] Simonnard Michel, *Linear Programming*, translated by William S. Jewell (Prentice Hall, Inc., Englewood Cliffs, N. J., 1966).

HIT PROBABILITY FOR A CHAIN-LIKE SERIES OF SHOTS—A SIMPLE FORMULA

T. U. Zahle

Division of Operations Research
Danish Defence Research Board

ABSTRACT

To find the hit probability for a series of correlated shots is, with conventional methods, a very tedious job even for the simplest statistical distributions. Experimental data, however, commonly show the shots lying chain-like on smooth curves about the target. For this case this paper introduces a new method of finding the hit probability without the use of correlation coefficients. The method is based on a transformation, where the target is transformed into a point and each bullet into an area the size of the target.

1. INTRODUCTION

In a military Operations Research organization one is often confronted with the task of determining the hit probability for a given weapon in a given combat situation. What is done is to make a number of experiments or attacks, each resulting in a shot picture showing how the shots have fallen around the target. The data can be obtained, for example, with cameras. From the shot pictures the probability (P) of scoring at least one hit can be estimated.

The simplest way to do this is to take the number of successful experiments and divide by the total number of experiments. This can always be done, but is not a very economical use of our data. Suppose in the extreme we have only one experiment (Figure 1):

FIGURE 1

To estimate the hit probability as either 1 or 0, according to whether we scored a hit or not, would clearly be inaccurate.

We can often get a better estimate if we can fit the data to a mathematical model. The point is that we can estimate the parameters of the model and then from these derive the hit probability, P, with better accuracy than by the simple approach.

Let us look at some often used models.

Model 1.

Assumption: The shots are independent. We can here estimate the probability distribution and the amount of probability mass (p) on the target. As the shots are independent we get

$$P=1-(1-p)^n,$$

where n is the number of shots.

Model 2.

Assumption: The aiming point is normally distributed around the target (systematic error), and the shots fall normally distributed around this aiming point (accidental error). The shots are correlated and our problem can here be solved by estimating the correlation coefficients from the data. These correlation coefficients are, in any case, difficult to extract from the data and when obtained lead to very tedious calculations, mostly based on infinite series which can be truncated at some point.

Model 3.

Assumption: There is a trend in the shots. More specifically the shots can be said to fall chain-like in smooth curves around the target. Here also the shots are correlated. The object of this paper is to find P for this model.

It will be proved that when the distance between shots is greater than the maximum dimension of the target, the correlation does not influence the hit probability, and we get $P=1-(1-p)^n$ as in the independent case.

When the distance between shots is smaller than the maximum width of the target, the formula

$$P=1-(1-kp)^n$$

can be used, where k is a reduction factor (to be defined later) between one and zero.

2. DEFINITION OF THE PROBLEM

We are given some experimental data in the form of shot-pictures for a particular situation, and we want to estimate the *probability of at least one hit in the target:*

1. We have only the shot-pictures as available data.

2. Nothing seems to indicate that any given shot in a series lies closer to the target than any other. The first shots for example have no special conditions compared to the other shots.

3. The shots on each picture are seen to arrange themselves *chain-like in smooth curves around the target.* (See Figure 2.) This is a situation which often arises for weapons firing a number of shots in a series. The exact meaning of "smooth curves" will be explained later.

FIGURE 2

"Shot picture" refers to the pattern described by the shots on a plane passed through the target perpendicular to the relative direction of fire.
"Target area" refers to the size of the target in this plane.

3. TRANSFORMATION

We set out by transforming the target into a point and the individual shots into plots each of whose area equals the apparent area of the target (Figure 3).

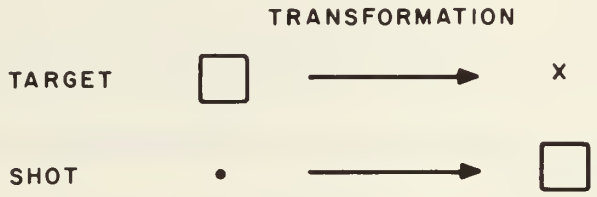


FIGURE 3

When a shot hits the target, the area representing the shot covers the point representing the target (Figure 4).

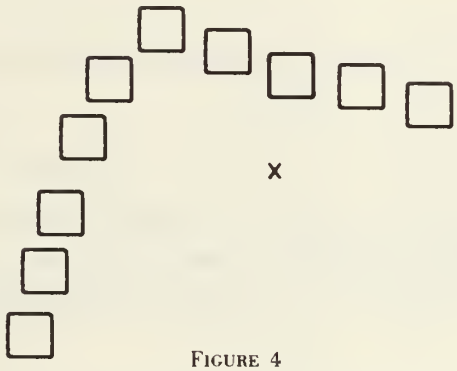


FIGURE 4

This transformation leaves the hit probability equal to the corresponding probability in the original situation.

4. RELEVANCE OF CORRELATION COEFFICIENTS (CASE 1)

Let us first look at a simple example where we have made the assumption that the shots are uniformly distributed over a given hitting area A, or that the chain moves in a smooth curve inside a given area A, in such a way that the area becomes evenly covered in the long run. This will later be referred to as random movement.

Further, we assume that the distance between two consecutive shots is larger than the maximum dimension of the target, so that consecutive plots in the transformed figure do not overlap each other. (See figure 5.)

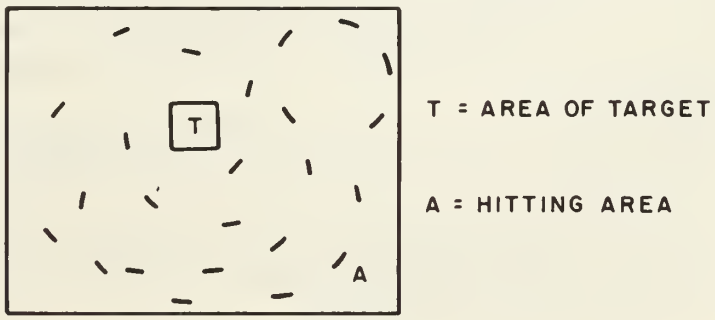


FIGURE 5

Consider Figures 6–8 "p" is defined as the hit probability for a single shot; in this case $p=T/A$.

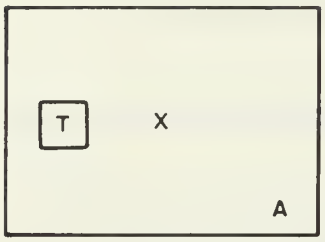


FIGURE 6

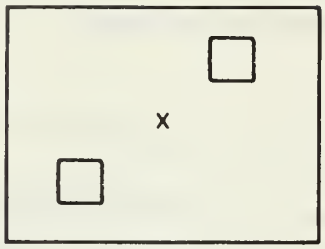


FIGURE 7

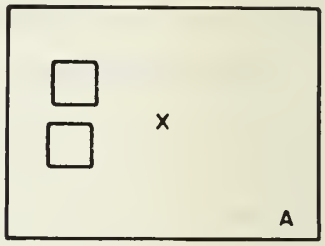


FIGURE 8

In Figure 6, only one shot has been fired, and as it falls with equal probability anywhere inside the area A, the hit probability, P, is

$$P=\frac{T}{A}=p.$$

In Figure 7, two independent shots have been fired, and the hit probability is

$$P=1-(1-p)^2=2p-p^2,$$

$$P\simeq 2p \qquad \text{for } p \text{ small.}$$

We see that P is a little less than $2p$. This follows from the law of diminishing returns (there is a chance that the second plot covers the first, hence the average area covered is less than $2T$).

In Figure 8, two chain-like shots have been fired and the hit probability is (in analogy with Figure 6)

$$P=\frac{2T}{A}=2p.$$

Because of our assumptions here the second plot cannot overlap the first, and there is no subtracted term.

We now realize that the correlation between the two shots on Figure 8 (i.e., when one knows where the first shot falls, one has much information about where the second is going to fall), is not necessary going to influence the hit probability in a negative way.

It is often said that correlated shots must give poorer results than independent shots, but this is only because one does not distinguish between different kinds of correlation.

What is usually meant by correlation is that a given new shot falls according to a normal distribution around the previous shot, and in this case it is clear that the plots have a larger probability of overlapping than when they are independent. This is why the assumption about the shots being placed along *smooth* curves has such important consequences (by this we have eliminated the tendency of consecutive plots to overlap at all).

In discussions on this subject one often encounters the following argument: "Well, but if the first shot falls far from the target, the second shot is going to miss too, and so the correlation must have a negative effect on the results." This argument does not hold, because in the case where the first shot falls close to the target we get a correspondingly greater chance of hitting it in the follwoing shot, and so the two tendencies might (indeed) cancel each other.

Do not think of it as the hit probability of the first shot plus the hit probability of the second, etc., because then inevitably we begin to operate with conditional probabilities. Think of it in this way: We place the two shots in a certain pattern, and then *throw the pattern as a whole on the area A.*

5. HIT PROBABILITY FOR A UNIFORM DISTRIBUTION OF SHOTS (CASE 1)

We have here the same assumptions as in section 4:

1) Shots uniformly distributed over a hitting area A.

2) When shots fall chain-like in smooth curves the distance between two consecutive shots is larger than the maximum dimension of the target.

If the shots were independent (Figure 9a), we know the hit probability would be

$$P = 1 - (1-p)^n.$$

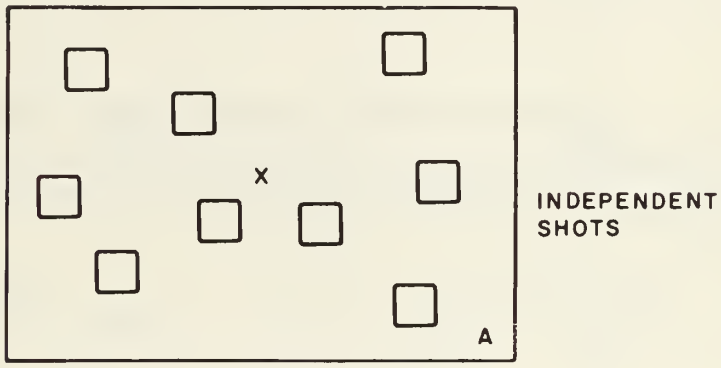


FIGURE 9A

This result could also be derived by considering how large an area the plots would cover on the average and dividing this area by A. As we need this line of reasoning for later analogy it will be explained shortly.

We define the "covered area" (C_i) as the area that on the average is covered by i plots.

Each new plot increases on the average the covered area with

$$\Delta C = T\left[\frac{\text{Area of } A \text{ not yet covered}}{\text{Total area } A}\right].$$

This means that if we, at a certain point, have covered an area C_i our next plot will on the average increase the covered area by

$$\Delta C_{i+1} = T\left[\frac{A - C_i}{A}\right] = T\left[1 - \frac{C_i}{A}\right].$$

We therefore get

$$\Delta C_1 = T \qquad C_1 = T = A\left[1-\left(1-\frac{T}{A}\right)\right]$$

$$\Delta C_2 = T\left[1-\left(1-\left(1-\frac{T}{A}\right)\right)\right] = T\left(1-\frac{T}{A}\right) \qquad C_2 = A\left[1-\left(1-\frac{T}{A}\right)\right]+T\left(1-\frac{T}{A}\right) = A\left[1-\left(1-\frac{T}{A}\right)^2\right]$$

$$\Delta C_3 = T\left[1-\left(1-\left(1-\frac{T}{A}\right)^2\right)\right] = T\left(1-\frac{T}{A}\right)^2 \qquad C_3 = A\left[1-\left(1-\frac{T}{A}\right)^2\right]+T\left(1-\frac{T}{A}\right)^2 = A\left[1-\left(1-\frac{T}{A}\right)^3\right]$$

$$\Delta C_n = T\left(1-\frac{T}{A}\right)^{n-1} \qquad C_n = A\left[1-\left(1-\frac{T}{A}\right)^n\right].$$

The probability of getting at least one hit for n shots is

$$P=\frac{C_n}{A}=1-\left(1-\frac{T}{A}\right)^n, \qquad \text{or}$$

$$\boxed{P=1-(1-p)^n},$$

where p is equal to T/A or the probability mass on target for a single shot.

For the sake of completeness it should be noted that for large n,

$$P \simeq 1-e^{-np}.$$

This is the result for independent shots. What is the difference when we have shots lying chain-like (Figure 9b) in smooth curves around the target?

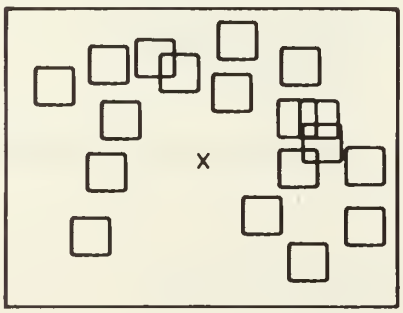


FIGURE 9B

The assumption *chain-like* implies that shots follow one another like elements in a chain. Therefore, from knowledge of previous shots we can pretty well predict where the next shot is going to fall. This influence, however, is going to weaken for the following shots, and after a certain point the influence is zero. That is, nothing can be said from the position of shot no. i about the position fo shot no. $(i+r)$. Said in another way: The chain has its different (not too near) portions placed independently of one another in A.

The assumption of *smooth* curves furthermore implies that nearby plots have no excessive tendency to overlap, i.e., no more probability of over-lapping than they would have in the independent case.

We therefore have the same situation as for independent shots in regard to covering area C_n with the difference that consecutive shots cannot overlap. The covered area for the chain-like case is therefore greater. From this it follows that the formula for independent shots is a lower limit approximation to the chain-like case.

It is furthermore a good approximation, because the change in covered area from the independent case where consecutive plots have a slight chance of overlapping to the chain case where they have no chance at all, is normally small.

The situation here resembles somewhat that of Koopman, in his article [4] which inspired this work, because we were puzzled to see that adjacent "observations" gave the same result as independent ones.

6. HIT PROBABILITY FOR OTHER DISTRIBUTIONS (CASE 2)

Suppose the chain moves in such a way that the distribution of shots becomes trapezoidal (Figure 10a). A cross-sectional slice of this bivariate density distribution is seen in Figure 10a.

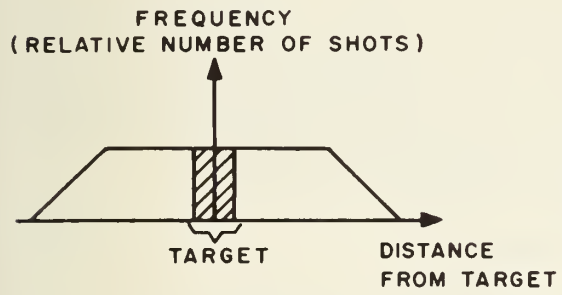


FIGURE 10A

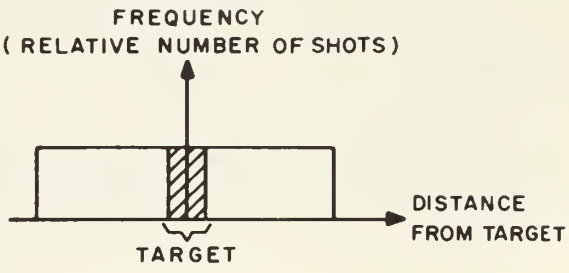


FIGURE 10B

When we compare this situation with the uniform distribution (Figure 10b) we see that the only difference is in how the distant shots fall. This difference, however, is irrelevant for hit probability because it is *after all unimportant whether shots which do not hit the target fall at great distance from the target or close to it.* Therefore the two situations must have the same hit probability, and we know that the hit probability for Figure 10b is

$$P=1-(1-p)^n.$$

The only relevant information is the probability mass on target (p) and the number of shots, but not the distribution.

The same reasoning applies to distributions other than the trapezoidal; e.g., when the chain moves in such a way that the density of shots becomes a normal distribution (Figures 11a, b).

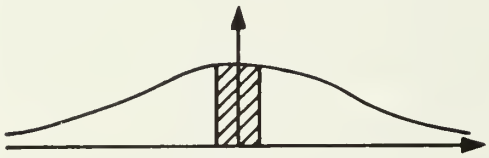


FIGURE 11A

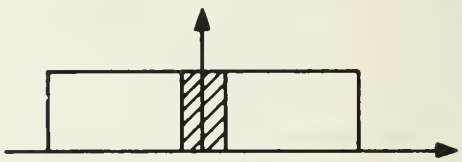


FIGURE 11B

Here again we can compare it with a uniform distribution which has the same probability mass on target, and we see that the total hit probability must be the same, i.e., $P=1-(1-p)^n$.

7. EXAMPLE

Say we are given the following three pictures (Figure 12) and we want to make an estimate of the hit probability:

FIGURE 12

We see that the assumptions of:

1. Chain-like shots in smooth curves, and
2. Greater distance than maximum width,

are fulfilled.

By placing the pictures one over the other we make an estimate of the probability distribution. Then we integrate the distribution over the target area, T, and get, say, $p=0.05$.

With nine shots the hit probability is:

$$P=1-(1-0.05)^9\approx 0.4.$$

8. HIT PROBABILITY WHEN SHOTS ARE CLOSER THAN THE MAXIMUM WIDTH OF TARGET (CASE 3)

When consecutive shots lie closer together than the maximum width of the target, we have a picture like the following (Figure 13):

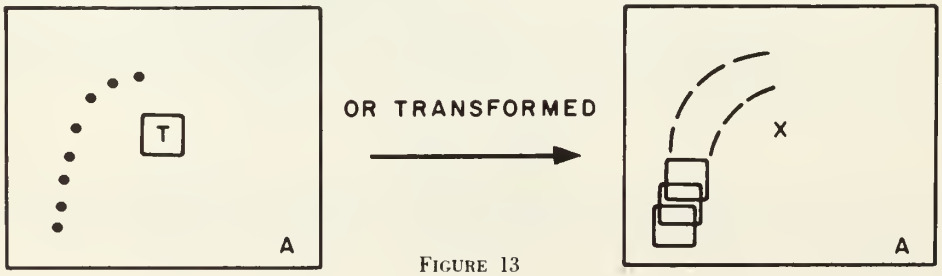


FIGURE 13

Here each new plot does not increase the chain with the area, T, but only with a reduced area (Tk). Taking this fact into account we can repeat all our previous arguments with each new plot having the size (Tk). This leads to the formula:

$$P = 1 - \left(1 - \frac{Tk}{A}\right)^n \qquad \text{or}$$

$$P = 1 - (1 - pk)^n,$$

where the reduction factor, k, is the average fraction of the area, T, which lies outside the immediately preceding plot.

The inherent overlap of consecutive shots is in nature quite different from the stochastic overlap which already is taken into account in the formula $P = 1 - (1 - p)^n$. This reduction obtained by the *inherent* overlap is of course directly due to the correlation, but has in this method the advantage of being handled in a very concrete way.

An estimate of k could be made as follows:

1) Obtain a lot of shot films.

2) Transform the shots into areas.

3) For each film, compute

$$k = \frac{\text{Total area covered}}{\text{No. of shots} \times T}.$$

4) Average the k's.

Alternatively, if the target area, T, is the same for each shot-picture, step (3) can be carried out at once for all the pictures, yielding

$$\hat{k} = \frac{\text{Total area covered on all pictures}}{\text{Total no. of shots on all pictures} \times T}.$$

It should be noted that the above procedure is only valid when the average distance between consecutive shots is the same, regardless of distance from target. What we are after is the inherent overlap when shots fall close to the target, so when we estimate k we must limit ourselves to the shots which have the same average distance as those over the target.

Another difficulty arises when the target size changes during an attack or from one attack to another, because this influences P as well as k. This difficulty, however, arises in other approaches as well. The easiest solution probably is to divide the attack into separate phases where p and k are constant and find p_i for each of these. We finally get

$$P = 1 - (1 - P_1)(1 - P_2) \ldots (1 - P_m),$$

where m is the number of phases.

If the target and the gun do not move in relation to one another k may be determined from the very design of the weapon used to fire the shots; e.g., suppose an automatic machine-gun were programmed so that its muzzle moved so as to fire successive shots in a figure "eight" at 3-inch intervals (Figure 14).

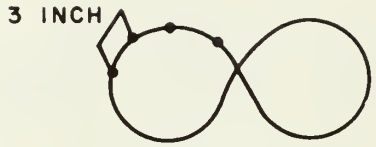


FIGURE 14

Given the target size and the distance to the target, it is possible to compute k.

9. DISCUSSION

The reaction of several military experts to this method has made it clear that it is difficult for a knowledgeable reader to accept the fact that we do not worry about correlation coefficients. The reaction is that, in one way or another, the problem is oversimplified and it is insisted upon that the correlation must show up somewhere in the calculations. However, the advantage of the method lies in the very fact that we arrive at the final hit probability without using correlation coefficents. The correlation coefficients are auxiliary variables which *can* lead us to the final result but they are *not necessarily* required in order to obtain the hit probabilities.

The constant, k, can in some respects be said to play the role of the correlation coefficient in other approaches, but the two parameters are by no means synonymous. In cases 1 and 2 the correlation has no influence on the hit probability, so it is in this respect irrelevant.

In the introduction we mentioned different mathematical models into which our data might fit. It is important to notice that while a considerable reduction in P is due to the correlation in the second model, no reduction in P necessarily arises due to the correlation in the third model. This stresses the importance of examining whether there is a trend in the shots and of fitting the data to the right model. It might be *misleading to talk nonspecifically about correlation.*

Besides the specific problem solved here, it is believed that the approach used, with the transformation of the bullets to plots of target size and the distinction between stochastic and inherent overlap, often can be used with advantage to get a grip on more complicated problems and lead to acceptable approximations.

10. CONCLUSION

We have described a method for finding the *probability of at least one hit* on the target, for series of shots which arrange themselves *chain-like in smooth curves* around the target. The method is very straightforward and leads to the simple formula

$$P = 1 - (1 - kp)^n.$$

11. ACKNOWLEDGMENT

I would like to thank my colleagues at the Division of Operations Research for useful discussions and, in particular, Mr. Torben Hansen, former head of the division, on whose initiative this work started.

12. REFERENCES

[1] Borenius, G., *Hit Probability in Salvos and Series* (Almqvist and Wiksells, Boktryckeri Aktiebolag, Uppsala, Sweden, 1963).

[2] Cunningham, L. B. C. and W. R. B. Hynd, "Random Processes in Problems of Air Warfare," J. Royal Statist. Soc., Supplement (1946).

[3] Fraser, D. A. S., "Generalized Hit Probabilities with a Gaussian Target II," Ann. Math. Statist. 24 (1953).

[4] Koopman, B. O., "The Theory of Search II, Target Detection," Operations Research *4*, 503–531 (1956), esp. the section "Random Search," pp. 519–521.

[5] Thompson, W. A., Jr. "The Diffused Target Concept," Operations Research *6*, 671–675 (1958).

[6] von Neumann, John, "Optimum Aiming at an imperfectly Located Target," appendix to "Optimum Spacing of Bombs or Shots in the Precence of Systematic Errors," by L. S. Dedrick and R. H. Kent, Ballistic Research Laboratory Report 241 (July 3, 1941); also published in pp. *v. Neumann's Collected Works*, edited by A. H. Taub (Pergamon Press, New York, 1962), Vol. 4, pp. 492–506.

[7] Wagepetersen, A., *Traeffesandsynlighted ved skydning med flere skud* (Artilleriskolen, København S., Danmark, 1964).

☆ U. S. GOVERNMENT PRINTING OFFICE : 1971 483-713/1

INFORMATION FOR CONTRIBUTORS

The NAVAL RESEARCH LOGISTICS QUARTERLY is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts and other items for publication should be sent to The Managing Editor, NAVAL RESEARCH LOGISTICS QUARTERLY, Office of Naval Research, Arlington, Va. 22217. Each manuscript which is considered to be suitable material for the QUARTERLY is sent to one or more referees.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Refereeing may be expedited if an extra copy of the manuscript is submitted with the original.

A short abstract (not over 400 words) should accompany each manuscript. This will appear at the head of the published paper in the QUARTERLY.

There is no authorization for compensation to authors for papers which have been accepted for publication. Authors will receive 250 reprints of their published papers.

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics, for possible publication in the NEWS AND MEMORANDA or NOTES sections of the QUARTERLY.

NAVAL RESEARCH
LOGISTICS
QUARTERLY

JUNE 1971
VOL. 18, NO.
NAVSO P-1278

CONTENTS

ARTICLES Page

Operations Analysis During the Underwater Search for *Scorpion* by H. R. Richardson and L. D. Stone 141

Optimum Sequential Search with Discrete Locations and Random Acceptance Errors by M. W. Smith and J. E. Walsh 159

Information in Two-Stage Programming under Uncertainty by D. P. Baron 169

Optimal Control of Multi-Channel Service Systems by M. J. Magazine 177

The Transportation Paradox by W. Szwarc 185

A Backward Recursive Technique for Optimal Sequential Sampling Plans by A. W. Wortham and E. B. Wilson 203

The Interface of Computer Science and Statistics by A. F. Goodman 215

Statistical Concepts in Computational Mathematics by M. L. Juncosa 231

Extended Iterative Weighted Least Squares: Estimation of a Linear Model in the Presence of Complications by A. F. Goodman 243

The Lower Bounded and Partial Upper Bounded Distribution Model by A. Charnes, Fred Glover and D. Klingman 277

Hit Probability for a Chain-Like Series of Shots—A Simple Formula by T. U. Zahle 283

OFFICE OF NAVAL RESEARCH
Arlington, Va. 22217